

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 04-293161

(43)Date of publication of application : 16.10.1992

(51)Int.Cl.

G06F 15/40

G06F 15/20

G06F 15/40

(21)Application number : 03-080547

(71)Applicant : HITACHI LTD

(22)Date of filing : 20.03.1991

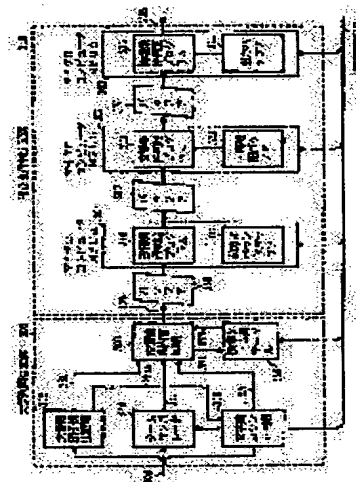
(72)Inventor : KAWAGUCHI HISAMITSU  
AKISAWA MITSURU  
KATO KANJI  
HATAKEYAMA ATSUSHI  
FUJISAWA HIROMICHI

## (54) METHOD AND DEVICE FOR RETRIEVING DOCUMENT

### (57)Abstract:

**PURPOSE:** To obtain the concrete method and device for retrieving a document, which can decide a composite condition of a neighborhood condition, a context condition, a logical condition, etc., in a full text search.

**CONSTITUTION:** By a device constituted of a character-string collating circuit 200 and a composite condition deciding circuit 300, a character-string collation of a document of a document data base and a designated retrieval word in a retrieval conditional expression is executed, and when the collation is executed, an identifier of the retrieval word collated with document discriminating information and collating information in the document are outputted as collating information. At the time of deciding a composite condition, a decision of a neighborhood condition is executed by checking the formation of a close distance condition between the retrieval words designated by the retrieval conditional expression, based on the collating information, the collating information and collating information of a result of decision are outputted as the whole collating information, and subsequently, a logical condition is decided by checking the common occurrence of the designated retrieval words in the same phrase, sentence, etc., based on the whole collating information, the whole collating information is outputted, and next, the logical condition is executed by checking a logical condition between the designated retrieval words, and the whole collating information obtained therein is outputted as final retrieval result information.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than

**THIS PAGE BLANK (USPTO)**

特開平4-293161

(43)公開日 平成4年(1992)10月16日

(51)Int.Cl.<sup>3</sup>

識別記号

庁内整理番号

F I

技術表示箇所

G 0 6 F 15/40

5 0 0 D 7056-5L

15/20

5 3 4 Z 6945-5L

15/40

5 1 0 Z 7056-5L

審査請求 未請求 請求項の数16(全 51 頁)

(21)出願番号

特願平3-80547

(22)出願日

平成3年(1991)3月20日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 川口 久光

東京都国分寺市東恋ヶ窪1丁目280番地

株式会社日立製作所中央研究所内

(72)発明者 秋沢 充

東京都国分寺市東恋ヶ窪1丁目280番地

株式会社日立製作所中央研究所内

(72)発明者 加藤 寛次

東京都国分寺市東恋ヶ窪1丁目280番地

株式会社日立製作所中央研究所内

(74)代理人 弁理士 笹岡 茂 (外1名)

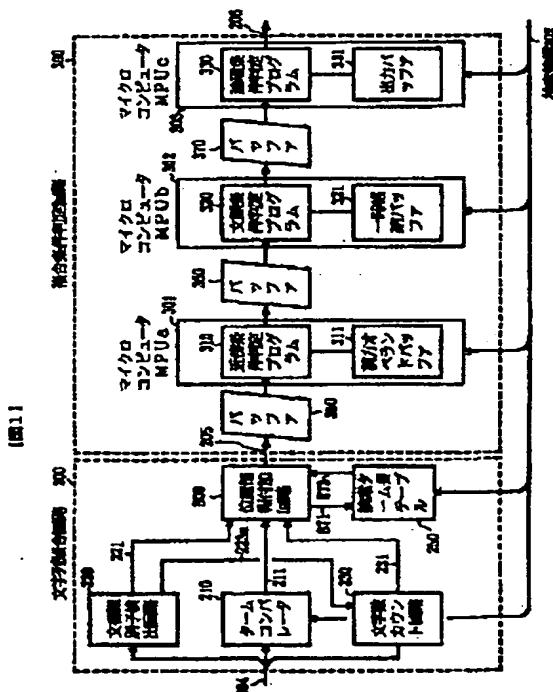
最終頁に続く

(54)【発明の名称】 文書検索方法および装置

(57)【要約】

【目的】 フルテキストサーチにおける近傍条件、文脈条件、論理条件等の複合条件の判定ができる具体的な文書検索方法および装置を提供する。

【構成】 文書データベースの文書と検索条件式中の指定検索語との文字列照合をして、照合があったとき、その文書を特定する文書識別情報と照合された検索語の識別子および該文書中における照合位置とを照合情報として出力する。複合条件判定では、前記照合情報に基づいて検索条件式に指定された検索語間の近接距離条件の成立を調べて近傍条件の判定を行い、前記照合情報と判定結果の照合情報とを全照合情報として出力し、次いで前記全照合情報に基づき同一の句、文等の内での指定検索語の共起を調べて文脈条件判定を行い、全照合情報を出力し、次いでこの全照合情報に基づき指定検索語間の論理条件調べて論理条件判定を行い、ここで得られた全照合情報を最終的な検索結果情報として出力する。



1

## 【特許請求の範囲】

【請求項1】 文字コードとして蓄積された文書データベースを対象として検索条件式中に指定された検索語を含む文書を検索する文書検索方法において、文書中に指定検索語が照合された場合、該文書の識別子を含む文書識別情報と照合された検索語の識別子および該文書中における照合位置とを照合情報として出力する文字列照合ステップと、該文字列照合ステップで出力された照合情報を基に前記検索条件式中に指定された検索語間の位置関係に関する検索条件を判定し、検索条件に合致したことを示す判定結果の照合情報を作成し検索結果として出力する複合条件判定ステップから構成されることを特徴とする文書検索方法。

【請求項2】 請求項1記載の文書検索方法における複合条件判定ステップとして、検索条件式中に指定された検索語間の近接距離条件について判定を行う近傍条件判定ステップを有することを特徴とする文書検索方法。

【請求項3】 請求項1記載の文書検索方法における複合条件判定ステップとして、検索条件式中に指定された検索語の同一句、同一文、あるいは同一段落内での共起条件について判定を行う文脈条件判定ステップを有することを特徴とする文書検索方法。

【請求項4】 請求項1記載の文書検索方法における複合条件判定ステップとして、検索条件式中に指定された検索語間の論理条件について判定を行う論理条件判定ステップを有することを特徴とする文書検索方法。

【請求項5】 請求項1記載の文書検索方法における複合条件判定処理ステップとして、検索条件式中に指定された検索語間の近接距離条件について判定を行う近傍条件判定ステップと、該検索条件式中に指定された該検索語の同一句、同一文、あるいは同一段落内での共起条件について判定を行う文脈条件判定ステップと、該検索条件式中に指定された該検索語間の論理条件について判定を行う論理条件判定ステップを有することを特徴とする文書検索方法。

【請求項6】 請求項2記載の文書検索方法における複合条件判定ステップとして、文書データが日本文の場合、検索条件式中に指定された検索語間の文字数で表した近接距離条件について判定を行う近傍条件判定ステップを有することを特徴とする文書検索方法。

【請求項7】 請求項2記載の文書検索方法における複合条件判定ステップとして、文書データが英文の場合、検索条件式中に指定された検索語間の語数で表された近接距離条件について判定を行う語間条件判定ステップを有することを特徴とする文書検索方法。

【請求項8】 請求項5記載の文書検索方法において、文字列照合ステップとして、文書中に指定検索語が照合された場合、該文書の識別子を含む文書識別子と照合された検索語の識別子および該文書中における照合検索語の先頭文字位置と末尾文字位置を照合情報として出力

2

し、文脈条件が指定され、文脈を識別する文字列が照合された場合、該文書の識別子と照合された文脈識別文字列の識別子および該文書中における該照合文脈識別文字列の先頭位置と末尾位置を照合情報として出力するとともに、複合条件判定ステップを構成する近傍条件判定ステップ、文脈条件判定ステップ、及び論理条件判定ステップにおいて、近傍条件判定ステップでは前記文字列照合ステップで出力された照合情報に基づいて検索条件式中に指定された検索語間の文字数で表した近接距離条件について判定を行い、条件に合致した前方に位置する検索語の先頭文字位置と後方に位置する検索語の末尾文字位置を判定結果の照合情報として、これを前記文字列照合ステップで出力された照合情報に付加して出力し、文脈条件判定ステップでは前記近傍条件判定ステップで出力された照合情報に基づいて該検索条件式中に指定された検索語の同一句、同一文、あるいは同一段落内での共起条件について判定を行い、条件に合致した前方に位置する文脈識別文字列の先頭文字位置と後方に位置する文脈識別文字列の末尾文字位置を照合情報として、これを前記近傍条件判定ステップで出力された照合情報に付加して出力し、論理条件判定ステップでは前記文脈条件判定ステップで出力された照合情報に基づいて該検索条件式中に指定された該検索語間の論理条件について判定を行い、条件に合致した文書単位の照合情報を最終的な検索結果情報として出力することを特徴とする文書検索方法。

【請求項9】 文字コードとして蓄積された文書データベースを対象として検索条件式中に指定された検索語を含む文書を検索する文書検索装置において、文書中に指定検索語が照合された場合、該文書の識別子を含む文書識別情報と照合された検索語の識別子および該文書中における照合位置とを照合情報として出力する文字列照合手段と、該文字列照合手段で出力された照合情報を基に前記検索条件式中に指定された検索語間の位置関係に関する検索条件を判定し、検索条件に合致したことを示す判定結果の照合情報を作成し検索結果として出力する複合条件判定手段から構成されることを特徴とする文書検索装置。

【請求項10】 請求項9記載の文書検索装置における複合条件判定手段として、検索条件式中に指定された検索語間の近接距離条件について判定を行う近傍条件判定手段を備えることを特徴とする文書検索装置。

【請求項11】 請求項9記載の文書検索装置における複合条件判定手段として、検索条件式中に指定された検索語の同一句、同一文、あるいは同一段落内での共起条件について判定を行う文脈条件判定手段を備えることを特徴とする文書検索装置。

【請求項12】 請求項9記載の文書検索装置における複合条件判定手段として、検索条件式中に指定された検索語間の論理条件について判定を行う論理条件判定手段

を備えることを特徴とする文書検索装置。

【請求項13】 請求項9記載の文書検索装置における複合条件判定処理手段として、検索条件式中に指定された検索語間の文字数で表した近接距離条件について判定を行う近傍条件判定手段と、該検索条件式中に指定された該検索語の同一句、同一文、あるいは同一段落内での共起条件について判定を行う文脈条件判定手段と、該検索条件式中に指定された該検索語間の論理条件について判定を行う論理条件判定手段を備えることを特徴とする文書検索装置。

【請求項14】 請求項10記載の文書検索装置における複合条件判定手段として、文書データが日本文の場合、検索条件式中に指定された検索語間の文字数で表した近接距離条件について判定を行う近傍条件判定手段を備えることを特徴とする文書検索装置。

【請求項15】 請求項10記載の文書検索装置における複合条件判定手段として、文書データが英文の場合、検索条件式中に指定された検索語間の語数で表された近接距離条件について判定を行う語間条件判定手段を有することを特徴とする文書検索装置。

【請求項16】 請求項13記載の文書検索装置において、前記文字列照合手段は、文書中に指定検索語が照合された場合、該文書の識別子と照合された検索語の識別子および該文書中における照合検索語の先頭文字位置と末尾文字位置を照合情報として出力し、文脈条件が指定され、文脈を識別する文字列が照合された場合、該文書の識別子と照合された文脈識別文字列の識別子および該文書中における該照合文脈識別文字列の先頭位置と末尾位置を照合情報として出力するよう構成され、前記複合条件判定手段を構成する前記近傍条件判定手段は、前記文字列照合手段で出力された照合情報に基づいて検索条件式中に指定された検索語間の文字数で表した近接距離条件について判定を行い、条件に合致した前方に位置する検索語の先頭文字位置と後方に位置する検索語の末尾文字位置を判定結果の照合情報として、これを前記文字列照合手段で出力された照合情報に付加して出力するよう構成され、前記複合条件判定手段を構成する前記文脈条件判定手段は、前記近傍条件判定手段で出力された照合情報に基づいて該検索条件式中に指定された検索語の同一句、同一文、あるいは同一段落内での共起条件について判定を行い、条件に合致した前方に位置する文脈識別文字列の先頭文字位置と後方に位置する文脈識別文字列の末尾文字位置を照合情報として、これを前記近傍条件判定手段で出力された照合情報に付加して出力するよう構成され、前記複合条件判定手段を構成する論理条件判定手段は、前記文脈条件判定手段で出力された照合情報に基づいて該検索条件式中に指定された該検索語間の論理条件について判定を行い、条件に合致した文書単位の照合情報を最終的な検索結果情報として出力するよう構成されてなることを特徴とする文書検索装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は情報処理システム、特に情報検索システムにおけるフルテキストサーチに係り、近傍条件、文脈条件、および論理条件などの複合条件判定処理を高速に実現する方法および装置に関するものである。テキストデータベース、ワードプロセッサ、および文書ファイリングシステムなどにおける検索に利用し得るものである。

10 【0002】

【従来の技術】 近年、文献情報や特許情報などの2次情報(書誌情報)のみならず、1次情報(原文)をも含む大規模データベース・サービスの重要性が増してきている。従来、データベースの情報検索では、シソーラスに基づいて統制されたキーワードや分類コード等の2次情報による検索が行われてきている。しかし、この方法では数十件から数百件までにしか絞り込めないため、検索者が最終段階で直接本文を読んで内容を確認しなければならないという効率上の問題がある。また、分類体系自体が年月と共に変化するため、常にキーワードや分類コードを更新しなければならないという問題も生じてくる。更に、キーワード付け(インデキシングと言う)には時間がかかるため新たな文書はバッチ処理によりかなりの量をまとめて登録する。そのため、検索する情報は常に一定期間の送れを持つという問題がある。

【0003】 これらの問題に対処する一つの方法として、検索者が自由な検索語に基づいて、文書の本文を直接参照して内容を検索できるフルテキストサーチシステムが考えられている。このようなフルテキストサーチシステムを実現するための文書検索装置がいくつか提案されている。その中の代表的な文書検索装置の構成を第2図に示し、その内容について説明する。(エル エー ホラー : "ハードウェア システムズ フォー テキスト インフォメーション リトリバー", エー シー エム, エス アイ ジー アイ アール, 第6回コンファレンス 1983年, L. A. Hollaar: "Hardware systems for Text Information Retrieval", ACM SIGIR6th Conference 1983) 文書検索装置1において、検索制御手段101は検索装置全体の制御とホストコンピュータとの通信を行う。すなわち、ホストコンピュータから送られてくる検索要求201を受け付けこれを解析し、文字列照合手段200と複合条件判定手段300へ検索情報202として送出する。また、検索制御手段101は記憶装置制御手段104を制御して、文字列記憶手段105に格納された文書データ204を文字列照合手段200へ読み出す。文字列照合手段200は文書データ204の中に検索要求で指示された検索語(以後、検索タームと呼ぶ)に合致するものがあるかどうかを調べ、もし該

当するものがあれば、該当文字列を識別する情報205を複合条件判定手段300へ出力する。複合条件判定手段300は該文字列識別情報205に対して、検索要求中に指示されたANDやORなどで構成される論理条件などが満足されるか否かを調べる。複合条件が満足された場合には、該当する文書の識別情報や文書内容を検索結果206としてホストコンピュータへ返送する。

【0004】本システムでは絞り込みを精度良く行うために、複合条件判定手段300の検索条件として論理条件の他に、英文を対象とした以下に示す条件が提案されている。

“A . n. B” (1-1)

“<A, B>n” (1-2)

“A AND B IN SENT” (1-3)

(1-1)式の“A . n. B”という条件式は、“A”と“B”という2つの検索タームがこの順序で現れ、かつこの2つの検索タームがn単語以内に近接して現れる文書を探し出すことを表す。

(1-2)式の“<A, B>n”という条件式は、“A”と“B”という2つの検索タームがその順序を問わずに、すなわち“A”が“B”の前に現れる場合、あるいは“B”が“A”の前に現れる場合のどちらであっても、これらの検索タームがn語以内に近接して現れる文書を探し出すことを表す。

(1-1)式や(1-2)式のように検索ターム間の近接の度合いを尺度とする検索条件を近傍条件と呼ぶことにする。

(1-3)式の“A AND B IN SENT”という条件式は、“A”と“B”という2つの検索タームがその順序を問わずに、同一の文(センテンス)に現れる文書を探し出すことを表す。

(1-3)式のように文や段落(パラグラフ)という同一文脈(フィールドとも呼ぶ)上における2つの検索タームの共起を判定する条件を文脈条件と呼ぶことにする。

【0005】このように本引用文献では、検索ターム間の距離的な結び付きや文脈的な結び付きを制約とする近傍条件および文脈条件などの検索条件が複合条件として提案されている。これらの条件を用いると単に論理条件を用いて検索するのに比べ、キーワード間の意味的な結び付きを加味して検索できることになるため、木目細かな検索が行えることになり、その結果精度の良い絞り込みが出来ることになる。しかしながら、本引用文献には近傍条件や文脈条件を実現する具体的な方法が記述されていない。また、フルテキストサーチでは文書データを直接サーチするため処理時間が膨大となる。そこで検索タームを高速に探索するためにタームコンパレータと呼ぶ文字列照合用ハードウェアが提案されている。この具体的な実現方法は、例えば特開昭60-105039に開示されている。このタームコンパレータでは、数MByte/sから数十MByte/sと高速に文字列照合を

行うことが可能である。しかしながら、これらのタームコンパレータには文字列照合手段200と同等の機能はなく、検索の絞り込みに重要な近傍条件、文脈条件、および論理条件などの複合条件判定機能は搭載されていない。さらに、複合条件判定手段300では文字列照合手段200で高速に照合処理された大量の検索ターム(以後、照合タームと呼ぶ)を文字列照合手段200の照合処理速度と同じく高速に判定処理しなければならない。これは、文字列照合手段200がいくらか高速に処理できたとしても、複合条件判定手段300の処理が遅いとシステムとしての検索速度が落ちてしまうため検索時間を短縮することができないからである。したがって、複合条件判定手段300としては近傍条件、文脈条件、および論理条件を高速判定処理できるものでなければならないということになる。

【0006】

【発明が解決しようとする課題】本発明の課題は、フルテキストサーチ特有の木目細かな絞り込みを可能とする、近傍条件判定、文脈条件判定、および論理条件判定などの複合条件の具体的な判定方法を提供するとともに、これらの組合せ処理をハードウェア化された文字列照合手段と同等の速度で行うことのできる複合条件判定方法を提供することである。

【0007】本発明で具体的に実現しようとする複合条件検索機能は以下の通りである。まず近傍条件としては、日本語の場合には検索タームの間に存在する文字数に上限あるいは下限を指定した字間距離条件検索を、英語の場合には検索ターム間の語数に上限あるいは下限を指定した語間条件検索などを実現する。字間距離条件の例としては、以下のようなものがある。

“文書[8C]検索” (2-1)

“文書[10c]検索” (2-2)

“文書[8c, 10c]検索” (2-3)

“文書<10c>検索” (2-4)

(2-1)式の“文書[8C]検索”という条件式は、“文書”と“検索”という2つの検索タームがこの順序で現われ、かつこの2つの検索タームの間に8文字以内の文字が挟まっている文書を探し出すことを表わす。したがって、第3図に示した例文の中では、①と②を検索することになる。

(2-2)式の“文書[10c]検索”という条件式は、“文書”と“検索”という2つの検索タームがその順序を問わずに、すなわち“文書”が“検索”の前に現われる場合、あるいは“検索”が“文書”の前に現われる場合のどちらであっても、これらの検索タームが10文字以内に近接して現われる文書を探し出すことを表わす。したがって、第3図に示した例文の中では、①と②と③を検索することになる。

(2-3)式の“文書[8c, 10c]検索”という条件式は、2つの検索タームがその順序を問わず、8文字以

上離れていて、かつ10文字以内に近接して現われる文書を探し出すことを表わす。したがって、第3図に示した例文の中では、②と③を検索することになる。

(2-4)式の“文書<10c>検索”という条件式は、“文書”と“検索”という2つの検索タームがその順序を問わず、10文字以上離れて現われる文書を探し出すことを表わす。したがって、第3図に示した例文の中では、③と④を検索することになる。

【0008】次に、語間距離条件の例としては以下のようなものがある。

“text [8W] search” (3-1)

“text [10w] search” (3-2)

“text [8w, 10w] search” (3-3)

“text<10w>search” (3-4)

(3-1)式の“text [8W] search”という条件式は、“text”と“search”という2つの検索タームがこの順序で現われ、かつこの2つの検索タームの間に8語(ワード)以下の単語が挟まっている文書を探し出すことを表わす。

(3-2)式の“text [10w] search”という条件式は、“text”と“search”という2つの検索タームがその順序を問わずに、すなわち“text”が“search”の前に現われる場合、あるいは“search”が“text”の前に現われる場合のどちらであっても、これらの検索タームが10語以内に近接して現われる文書を探し出すことを表わす。

(3-3)式の“text [8w, 10w] search”という条件式は、“search”と“text”という2つの検索タームがその順序を問わず、8語以上離れていて、かつ10語以内に近接して現われる文書を探し出すことを表わす。

(3-4)式の“text<10w>search”という条件式は“text”と“search”という2つの検索タームがその順序を問わず、10語以上離れて現われる文書を探し出すことを表わす。以上が近傍条件としての課題となる。

【0009】次に文脈条件検索としては、日本語および英語とも次のようなものがある。

“文書 [ P ] 検索”。

“text [ P ] search” (4-1)

“文書 [ p ] 検索”。

“text [ p ] search” (4-2)

“文書 [ S ] 検索”。

“text [ S ] search” (4-3)

“文書 [ s ] 検索”。

“text [ s ] search” (4-4)

“文書 [ PH ] 検索”。

“text [ PH ] search” (4-5)

“文書 [ ph ] 検索”。

“text [ ph ] search” (4-6)

【0010】以下、日本語の例で説明する。

(4-1)式の“文書 [ P ] 検索”という条件式は、“文書”と“検索”という2つの検索タームがこの順序で、同一の段落(パラグラフ)に現われる文書を探し出すことを表わす。

(4-2)式の“文書 [ p ] 検索”という条件式は、“文書”と“検索”という2つの検索タームが順序を問わずに、同一の段落に現われる文書を探し出すことを表わす。

10 (4-3)式の“文書 [ S ] 検索”という条件式は、“文書”と“検索”という2つの検索タームがこの順序で、同一の文(センテンス)に現われる文書を探し出すことを表わす。

(4-4)式の“文書 [ s ] 検索”という条件式は、“文書”と“検索”という2つの検索タームが順序を問わずに、同一の文に現われる文書を探し出すことを表わす。

(4-5)式の“文書 [ PH ] 検索”という条件式は、“文書”と“検索”という2つの検索タームがこの順序で、同一の句(フレーズ)に現われる文書を探し出すことを表わす。日本語の場合、句とは“、”、“.”、“”、および“。”で区切られた文章を言う。英語の場合は“,”および“.”で区切られた文章ということになる。

(4-6)式の“文書 [ ph ] 検索”という条件式は、“文書”と“検索”という2つの検索タームが順序を問わずに、同一の句に現われる文書を探し出すことを表わす。以上が文脈条件としての課題となる。

30 【0011】最後に論理条件としては日本語および英語とも次のようなものがある。

“文書 [ AND ] 検索”。

“text [ AND ] search” (5-1)

“文書 [ OR ] 検索”。

“text [ OR ] search” (5-2)

“文書 [ NOT ] 検索”。

“text [ NOT ] search” (5-3)

【0012】以下、日本語の例で説明する。

(5-1)式の“文書 [ AND ] 検索”という条件式は、“文書”と“検索”という2つの検索タームが同時に現われる文書を探し出すことを表わす。

(5-2)式の“文書 [ OR ] 検索”という条件式は、“文書”あるいは“検索”という検索タームが現われる文書を探し出すことを表わす。

(5-3)式の“文書 [ NOT ] 検索”という条件式は、“文書”という検索タームが現われて、かつ“検索”という検索タームが現われない文書を探し出すことを表わす。以上が論理条件としての課題となる。

50 【0013】これらの課題をまとめると、本発明の課題はフルテキストサーチ特有の木目細かな絞り込みを可能とする、近傍条件判定、文脈条件判定、および論理条件

判定などの複合条件の具体的な判定方法を提供するとともに、これらの組合せ処理をハードウェア化された文字列照合手段と同等の速度で行うことの出来る複合条件判定方法を提供することである。

#### 【0014】

【課題を解決するための手段】これらの課題を解決するために、本発明の方法は、文字列照合ステップと複合条件判定ステップを備えている。文字列照合ステップにおいては、文書中に指定された検索タームが照合された場合、該文書の識別子である文書識別子と、照合された検索ターム、すなわち、照合タームの識別子および該文書中における照合タームの先頭文字位置と末尾文字位置を照合情報として出力し、文脈条件が指定され、文脈を識別する文字列が照合された場合、該文書の識別子と照合された文脈識別文字列の識別子および該文書中における該照合文脈識別文字列の先頭文字位置と末尾文字位置を照合情報として出力する。

【0015】複合条件判定ステップは、近傍条件判定ステップ、文脈条件判定ステップまたは論理条件判定ステップ、または、これら各ステップの組み合わせからなっている。近傍条件判定ステップにおいては、前記文字列照合ステップで出力された照合情報に基づいて検索条件式に指定された検索ターム間の文字数で表した近接距離条件について判定を行い、条件に合致した前方に位置する検索タームの先頭文字位置と後方に位置する検索タームの末尾文字位置を判定結果の照合情報として、これを前記文字列照合ステップで出力された照合情報に付加して出力する。

【0016】文脈条件判定ステップにおいては、検索条件式中に近傍条件が含まれている場合には、前記近傍条件判定ステップで出力された照合情報に基づいて該検索条件式中に指定された検索タームの同一句、同一文、あるいは同一段落内での共起条件について判定を行い、条件に合致した前方に位置する文脈識別文字列の先頭文字位置と後方に位置する文脈識別文字列の末尾文字位置を照合情報として、これを前記近傍条件判定ステップで出力された照合情報に付加して出力する。検索条件式中に近傍条件が含まれていない場合には、前記文字列照合ステップで出力された照合情報に基づいて該検索条件式中に指定された検索タームの同一句、同一文、あるいは同一段落内での共起条件について判定を行い、条件に合致した前方に位置する文脈識別文字列の先頭文字位置と後方に位置する文脈識別文字列の末尾文字位置を照合情報として、これを前記文字列照合ステップで出力された照合情報に付加して出力する。

【0017】論理条件判定ステップにおいては、検索条件式中に近傍条件が含まれている場合には前記近傍条件判定ステップで出力された照合情報に基づき、検索条件式中に近傍条件および文脈条件が含まれている場合または文脈条件が含まれている場合には前記文脈条件判定ス

テップで出力された照合情報に基づき、そして、検索条件式中に論理条件のみが含まれている場合には前記文字列照合ステップで出力された照合情報に基づいて、該検索条件式中に指定された該検索ターム間の論理条件について判定を行い、条件に合致した文書単位の照合情報を前段ステップで出力された照合情報に付加して最終的な検索結果情報として出力する。

【0018】また、本発明の装置の一つは次のように文字列照合手段と複合条件判定手段で構成される。文字列照合手段は、文書中に指定された検索タームが照合された場合、該文書の識別子である文書識別子と、照合された検索ターム、すなわち、照合タームの識別子および該文書中における照合タームの先頭文字位置と末尾文字位置を照合情報として出力し、文脈条件が指定され、文脈を識別する文字列が照合された場合、該文書の識別子と照合された文脈識別文字列の識別子および該文書中における該照合文脈識別文字列の先頭文字位置と末尾文字位置を照合情報として出力する。

【0019】複合条件判定手段は、近傍条件判定手段、文脈条件判定手段、および論理条件判定手段から構成される。近傍条件判定手段は、前記文字列照合手段で出力された照合情報に基づいて検索条件式に指定された検索ターム間の文字数で表した近接距離条件について判定を行い、条件に合致した前方に位置する検索タームの先頭文字位置と後方に位置する検索タームの末尾文字位置を判定結果の照合情報として、これを前記文字列照合手段で出力された照合情報に付加して出力する。文脈条件判定手段は、前記近傍条件判定手段で出力された照合情報に基づいて該検索条件式中に指定された検索タームの同一句、同一文、あるいは同一段落内での共起条件について判定を行い、条件に合致した前方に位置する文脈識別文字列の先頭文字位置と後方に位置する文脈識別文字列の末尾文字位置を照合情報として、これを前記近傍条件判定手段で出力された照合情報に付加して出力する。論理条件判定手段は、前記文脈条件判定手段で出力された照合情報に基づいて該検索条件式中に指定された該検索ターム間の論理条件について判定を行い、条件に合致した文書単位の照合情報を最終的な検索結果情報として出力する。

#### 【0020】

【作用】文字列照合において、文書識別子と、照合タームの識別子および文書中における照合タームの先頭文字位置と末尾文字位置とが照合情報として出力され、また、文脈条件が指定された際に、文脈識別文字列の識別子および文書中における該照合文脈識別文字列の先頭文字位置と末尾文字位置とが照合情報として出力されるので、近傍条件の判定は、検索条件式中の検索タームについては前記照合タームの識別子との一致をみることにより判定し、検索条件式中の字間距離条件については一致をみた各照合タームの先頭文字位置と末尾文字位置とを



比較判定することにより行われる。文脈条件の判定は、検索条件式中の検索タームについては、文字列照合、近傍条件の判定で得られた照合タームの識別子との一致をみることににより判定し、検索条件式中の検索タームが共起する範囲についての条件については、相前後する文脈識別文字列の識別子の位置と一致をみた各照合タームの先頭文字位置と末尾文字位置とを比較することにより、相前後する文脈識別文字列の識別子の位置の間に一致をみた各照合タームが共起することを判定することにより行われる。論理条件については、検索条件式中の検索タームについては、文字列照合、近傍条件の判定、文脈条件の判定で得られた照合タームの識別子との一致をみることににより判定し、一致をみた照合タームの識別子が検索条件式中の論理条件を満たしているか否かを判定することにより行われる。

【0021】そして、文字列照合手段および複合条件判定手段を構成することにより、上記の近傍条件、文脈条件、および論理条件などの複合条件判定を一貫して実現することができるためフルテキストサーチ特有の木目細かな検索が可能となる。さらに、例えば3つのマイクロコンピュータで各々、近傍条件判定処理、文脈条件判定処理、および論理条件判定処理を実行させることにより、各処理間で同期を取らなくとも処理動作させることが可能となる。すなわち、これらのマイクロコンピュータではそれぞれの入力バッファに照合情報が格納されるとこれに応じて条件判定処理を始めるというパイプライン処理を行うことが可能となり、高速な複合条件判定処理を実現することができる。

#### 【0022】

【実施例】最初に、本発明の方法および装置の原理について説明する。文字列照合手段において、まず文書データが入力された際、文書の先頭に格納された文書識別子が検出され照合結果として出力される。次に文書中に指定された検索タームが照合された場合、照合された照合タームの識別子と、該文書中における照合タームの照合位置として照合タームの先頭文字位置と末尾文字位置が照合情報として出力される。すなわち、1文書における照合情報としては、図13に示すように先頭に文書識別子があり、その次から検索タームの照合情報が来る構成となる。以上の処理が文書毎に全ての文書データを読み込み終えるまで繰り返し行われる。

【0023】具体的な検索タームの照合方法について図4を用いて説明する。例えば文字列照合手段に、検索ターム“文書”が設定され、“...文書理解を用いた検索システムである...”という文書が入力されたことを想定する。この場合の文字列照合手段の出力として得られる文書識別情報および照合ターム識別情報は(6-1)および(6-2)のように表す。

(D1, 0, 0) (6-1)

(T1, Xs, Xe) (6-2)

文書識別情報(6-1)においてD1は文書識別子を表し、これに続く2項は定数0(ゼロ)である。照合ターム識別情報(6-2)において、T1は検索タームの識別子(以後、照合ターム識別子と呼ぶ)を表し、Xsは文書中で探索された照合タームの先頭文字位置を、Xeは同様に末尾文字位置を表わす。図4の例では、文書識別情報は(D1, 0, 0)となり、“文書”の照合ターム識別情報は(T1, 31, 32)となる。

【0024】次に複合条件判定手段では、以下のような判定処理が行われる。まず、近傍条件判定手段では前記文字列照合手段で出力された照合情報に基づいて、検索ターム間の文字数で表した近接距離条件について判定が行われる。すなわち、検索条件式に指定された前方に位置する検索タームの末尾文字位置と後方に位置する検索タームの先頭文字位置との文字距離を算出し、この文字距離が近傍条件に指定された距離および順序を満たしているか否かの判定が行われる。近接距離条件が成立した場合には判定結果として、条件に合致した前方に位置する検索タームの先頭文字位置と後方に位置する検索タームの末尾文字位置を照合情報として、これを前記文字列照合手段で出力された照合情報に付加して出力する。具体的な近傍条件処理例を図5を用いて説明する。例えば“文書”と“理解”がこの順序で現れ、かつ4文字以内に近接する文書を検索するという近傍条件“文書[4C]理解”が設定され、文字列照合手段に文書“...文書理解を用いた検索システムである...”

が入力されたことを想定する。まず、検索タームとして“文書”と“理解”が文字列照合手段に設定される。文書が入力されると、この2つの検索タームについて文字列照合処理が実行され、以下の文書識別情報と照合ターム識別情報が得られる。

(D1, 0, 0) (6-3)

(T1, 31, 32) (6-4)

(T2, 33, 34) (6-5)

(6-3)は文書識別情報、(6-4)は“文書”の照合ターム識別情報、および(6-5)は“理解”の照合ターム識別情報である。次に、これらの情報に基づいて近傍条件“文書[4C]理解”について処理が行われる。

【0025】本例では、条件に合致した前方に位置する検索ターム“文書”の末尾文字位置である32と、条件に合致した後方に位置する検索ターム“理解”の先頭文字位置である33から文字距離は0((33-32)-1)=1-1=0)であることが算出でき、指定された4文字より小さいため、本例における近傍条件“文書[4C]理解”は成立していると判定される。最後に判定結果として、本近傍条件の識別子PIDをP1とし、条件に合致した前方に位置する検索ターム“文書”の先頭文字位置である31をXsに、後方に位置する検索タ

ーム“理解”の末尾文字位置である34をXeとした照合情報(6-6)(以後、近傍条件識別情報と呼ぶ。文脈条件の場合には文脈条件識別情報、論理条件の場合には論理条件識別情報と呼び、これらの情報を総称して複合条件識別情報と呼ぶ。)を、以下のように文字列照合手段で出力された照合情報に(6-3)、(6-4)、および(6-5)に付加して出力する。

(D1, 0, 0) (6-3)  
(T1, 31, 32) (6-4)  
(T2, 33, 34) (6-5)  
(P1, 31, 34) (6-6)

すなわち、近傍条件の判定結果を(PID, Xs, Xe)という検索タームの照合情報と同じ形式で照合ターム識別情報に付加する形で出力する。

【0026】次に文脈条件判定手段では前記近傍条件判定手段で出力された照合情報に基づいて検索条件式中に指定された検索タームの同一句、同一文、あるいは同一段落内での共起条件について判定を行う。共起条件判定では、条件に指定された文脈識別文字列の先頭文字位置から次の文脈識別文字列の末尾文字位置までの文脈範囲内に、2つの検索タームが条件中に指定された順序で現れているかどうかの判定を行う。共起条件が成立した場合には、判定結果として本文脈条件の識別子と条件に指定された前方に位置する文脈識別文字列の先頭文字位置と、後方に位置する文脈識別文字列の末尾文字位置を照合情報として、これを前記近傍条件判定手段で出力された照合情報に付加して出力する。

【0027】具体的な文脈条件の判定処理の例を図6を用いて説明する。ここでは、“文書”と“理解”がこの順序で現れ、かつ同一文内に共起する文脈条件“文書[S]検索”が設定され、文字列照合手段に“...文書理解を用いた検索システムである...”という文書(文書識別子=1)が入力されたことを想定する。まず、検索タームとして“文書”と“検索”が、さらに文脈条件が指定されているので、文脈を識別するための文字列“.”が文字列照合手段に設定される。文字列照合手段では、この3つの検索タームが照合され、図6に示すように

(D1, 0, 0) (6-7)  
(S1, 30, 30) (6-8)  
(T1, 31, 32) (6-9)  
(T3, 33, 34) (6-10)  
(S1, 48, 48) (6-11)

が出力される。照合ターム識別情報(6-8)および(6-11)において、S1は文脈を識別する文字列“.”の識別子を表している。

【0028】これらの照合情報は、近傍条件判定手段に送られるが、本例の場合では近傍条件が設定されていないため、近傍条件判定手段からは入力した照合情報が以下のようにそのまま出力される。

(D1, 0, 0) (6-7)  
(S1, 30, 30) (6-8)  
(T1, 31, 32) (6-9)  
(T3, 33, 34) (6-10)  
(S1, 48, 48) (6-11)

(6-7)は文書識別情報、(6-9)は“文書”の、(6-10)は“検索”の照合ターム識別情報の、(6-8)と(6-11)は“.”の照合ターム識別情報である。

10 【0029】次に、これらの照合情報に基づいて文脈条件“文書[S]検索”に関する共起条件判定が行われる。本例では、文脈すなわち文(センテンス)の範囲は、条件に指定された文脈識別文字列“.”の先頭文字位置である30文字目から次の文脈識別文字列“.”の末尾文字位置である48文字目までとして、すなわち(6-8)から(6-11)までの範囲で表される。本例の場合、この文脈範囲内に、条件に指定された検索ターム“文書”および“検索”がこの順序で含まれているため、“文書[S]検索”が成立していると判定される。ここで、“文書”と“検索”の順序関係は“文書”の末尾文字位置(32)と“検索”の先頭文字位置(40)を比較して判定される。すなわち、“文書”の末尾文字位置(32)が“検索”の先頭文字位置(40)より小さい(32<40)ので、“文書”の方が“検索”より前に位置していると判定できる。

30 【0030】最後に判定結果として、本文脈条件の識別子CIDをC1とし、条件に指定された前方に位置する文脈識別文字列“.”の先頭文字位置である30(先頭位置情報)と、後方に位置する文脈識別文字列“.”の末尾文字位置である48(末尾位置情報)を照合情報(6-12)(以後、文脈条件識別情報と呼ぶ)として、これを前記近傍条件判定手段で出力された照合情報に付加して以下のように出力する。

(D1, 0, 0) (6-7)  
(S1, 30, 30) (6-8)  
(T1, 31, 32) (6-9)  
(T3, 33, 34) (6-10)  
(S1, 48, 48) (6-11)  
(C1, 30, 48) (6-12)

40 【0031】最後に論理条件判定手段では前記文脈条件判定手段で出力された照合情報に基づいて該検索条件式中に指定された検索ターム間の論理条件について判定を行い、条件に指定された文書単位の照合情報を最終的な検索結果情報として出力する。具体的な論理条件の判定処理の例を図7を用いて説明する。例えば、“文書”と“検索”という2つの検索タームが同一文書中に現れる論理条件“文書[AND]検索”が設定され、文字列照合手段に文書

50 “...文書理解を用いた検索システムである...”

が入力されたことを想定する。まず、検索タームとして“文書”と“検索”が文字列照合手段に設定され、本図の例の文書が入力されると、文字列照合手段より以下の照合ターム識別情報が得られる。

(D0, 0, 0) (6-13)

(T1, 31, 32) (6-14)

(T3, 39, 40) (6-15)

(6-13)は文書識別情報、(6-14)は“文書”の、(6-15)は“検索”の照合ターム識別情報である。もし、近傍条件および文脈条件が設定されていない場合には、近傍条件判定手段および文脈条件判定手段を経由するかたちで、これらの情報はそのまま論理条件判定手段入力され、論理条件“文書[AND]検索”についての論理条件判定が行われる。

【0032】本例では論理条件判定手段において、“文書”と“検索”の照合タームが1文書内に同時に存在することを調べ、論理条件“文書[AND]検索”が成立するものと判定を下す。そして、本条件を満足した文書識別情報(6-13)と、本判定結果として本論理条件の識別子L1DをL1とし、該当文書の先頭文字位置と末尾文字位置とを照合情報(6-16)(以後、論理条件識別情報と呼ぶ)として、これを前記文脈条件判定手段の出力に付加して、以下のように出力する。

(D0, 0, 0) (6-13)

(T1, 31, 32) (6-14)

(T3, 39, 40) (6-15)

(L1, 0, 99) (6-16)

【0033】以上のように、検索条件式中に指定された検索ターム間の文字数で表した近接距離条件について判定を行う近傍条件判定手段と、検索条件式中に指定された検索タームの同一句、同一文、あるいは同一段落内での共起条件について判定を行う文脈条件判定手段と、検索条件式中に指定された検索ターム間の論理条件について判定を行う論理条件判定手段からなる複合条件判定手段を用いることにより、フルテキストサーチ特有の木目細かな絞り込み検索が可能となる。さらに各複合条件判定を構成する近傍条件判定処理、文脈条件判定処理、および論理条件判定処理の入出力情報形式が全く同じになっているため、これらを分散してパイプライン処理することにより高速な複合条件判定処理が可能となる。

【0034】次に、本発明における第1の実施例について図1を用いて説明する。本実施例は文字列照合回路200と複合条件判定回路300から構成されている。文字列照合回路200では検索制御手段101(図2)から送られてくる検索対象の検索タームと記憶装置制御手段104(図2)の制御の基に文字列記憶手段105(図2)から読み出される文書データ204とを照合し、照合されたものがあれば照合ターム情報を照合結果205として複合条件判定回路300へ送る。複合条件判定回路300では検索制御手段101から送られてく

る複合条件を基に文字列照合回路200より出力される照合ターム識別情報に関して複合条件が満たされるか否かを判定し、条件に合致する場合には該当する照合ターム識別情報と複合条件識別情報を文書単位に判定結果206として出力する。

【0035】まず、文字列照合回路200について詳細に説明する。文字列照合回路200は、タームコンパレータ210、文書識別子検出回路220、文字数カウント回路230、位置情報付加回路800、および検索ターム長テーブル250から構成されている。タームコンパレータ210では、指定された検索タームと送られてくる文書データ204との照合を行い、照合されたものがあればその検索タームの識別子である照合ターム識別子211(正の整数値データ)を、照合されたものがない場合には0(ゼロ)を位置情報付加回路800へ送出する。すなわち、照合ターム識別子211の値として、0の場合は無効データであり、正の整数の場合は有効データであり照合タームの識別子を表す。このタームコンパレータ210としては、特開昭60-105039に開示されているものを用いることができる。文書識別子検出回路220は、第8図に示すようにレジスタ224~228と229a、コンパレータ223、セレクト229から構成されている。レジスタ224には文書単位に文書データの先頭に付与されているトップオブテキストコード(TOT)が、レジスタ229aには0(ゼロ)が初期設定されている。8ビット単位に送られてくる文書データ204は、レジスタ225~レジスタ228から構成される4段のシフトレジスタ228bに次々と入力される。コンパレータ223では、この最終段出力228aについてレジスタ224に格納されているTOTと等しいか否かの比較を行う。レジスタ225~レジスタ228では各出力を32ビットの出力222としてセレクト229に送り、最終段出力228aがTOTの場合には、コンパレータ223よりセレクト信号223aが送られ、32ビットの出力222を文書識別子221として選択し、位置情報付加回路800へ送出する。また、セレクト信号223aが送られない間、セレクト229ではレジスタ229aに格納された0を選択する。すなわち、文書データの中から、文書単位に文書データの先頭に付与されているTOTが検出され、これに続いて格納されている32ビットの文書識別子221が位置情報付加回路800へ送出され、TOTが検出されない間は0の文書識別子221が送出されることになる。

【0036】文字数カウント回路230では、送られてくる文書データ204に対して、各文書毎に先頭からの8ビットの文字コード数をカウントし、1文字が2バイトで構成される文字数へ変換し、これを位置情報付加回路800へ送出する。文書毎に文字コードカウント値をリセットするのに、文書識別子検出回路220から出力

されるTOT検出信号223aが用いられる。検索ターム長テーブル250には、図9に示すように照合ターム識別子211をアドレスとするスロットに該当検索タームの長さが格納されており、位置情報付加回路800より照合ターム識別子を受け取り、位置情報付加回路800へ照合ターム識別子に対応した検索ターム長873を返送する。本図では、照合ターム識別子が1の検索ターム“理解”のターム長である2と、照合ターム識別子が2の検索ターム“システム”のターム長である4という情報が設定されている。従って、例えば“検索”に対応する照合ターム識別子211として1を受け取ると検索ターム長として2を送り返すことになる。

【0037】位置情報付加回路800は、図10に示すようにレジスタ810～816、ORゲート880～881、セクタ820～822、減算器830、および加算器831から構成されている。本回路の初期設定として、レジスタ813、レジスタ814、およびレジスタ815には0（ゼロ）が設定されており、各々セクタ820、セクタ821、およびセクタ822に0を出力している。また、レジスタ816には1が設定されており、加算器831に1を出力している。セクタ820～822ではセレクト信号890、891の両方が0の場合はZポートを選択する。すなわち各セクタは、レジスタ813、レジスタ814、およびレジスタ815を各々選択することになり、セクタ820～822の出力として0が照合情報205として出力されることになる。また、セレクト信号890が1でセレクト信号891が0の場合はXポートを、セレクト信号890が0でセレクト信号891が1の場合にはYポートを選択する。レジスタ810には文書識別子検出回路220から文書識別子221が送られる度に文書識別子221が格納されると共に、セクタ820およびORゲート880へ出力される。ここで、文書識別子221が0の場合は文書の識別子以外が検出されたことを示している。ORゲート880ではレジスタ810に文書識別子221が格納された際、文書識別子221の各ビット間の論理和を取り演算結果であるセレクト信号890がセクタ820～822に送られる。

【0038】したがって、レジスタ810に文書識別子221が格納された場合には文書識別子221の値は0（ゼロ）ではないため、ORゲート880の演算結果890が1になる。これを受けてセクタ820～822に1が出力されるため各セクタではXポートが選択される。照合情報205には、識別子としてレジスタ810に格納されている文書識別子221が、先頭位置情報としてレジスタ814に格納されている0（ゼロ）が、末尾位置情報としてレジスタ813に格納されている0（ゼロ）が文書識別情報として出力されることになる。従って、文書識別子は図11に示すように32ビットの文書識別子と、先頭位置情報が0でかつ末尾位置情報も

0である32ビットの0の固定値から構成されることになる。

【0039】レジスタ811にはタームコンパレータ210から照合ターム識別子211が送られる度に照合ターム識別子211が格納されると共に、この照合ターム識別子211はセクタ820、ORゲート881、および検索ターム長テーブル250へ出力される。さらに検索ターム長テーブル250から照合ターム識別子211に応じて検索ターム長873が読み出され減算器830に出力される。ここで、照合ターム識別子211が0の場合は検索タームが照合されていないことを示している。また、レジスタ812には文字数カウント回路230から文字数カウント231が送られる度に、文字数カウント231は照合タームの末尾位置情報812aとして格納されると共に、減算器830およびセクタ822へ出力され、さらに減算器830で末尾位置情報812aから検索ターム長873を引き、さらに加算器831で1加えられた照合タームの先頭位置情報831aがセクタ821に出力される。ORゲート881ではレジスタ811に照合ターム識別子211が格納された際、照合ターム識別子211の各ビット間の論理和を取り演算結果であるセレクト信号891がセクタ820～822に送られる。

【0040】したがって、レジスタ811に照合ターム識別子211が格納された場合には、照合ターム識別子211の値は0でないため、ORゲート881の出力が1となる。これを受けてセレクト信号891が1としてセクタ820～822に送られるため各セクタではYポートが選択され、照合情報205には、識別子としてレジスタ811に格納されている照合ターム識別子211と、先頭位置情報として加算器831から出力されている先頭位置情報831aと、末尾位置情報としてレジスタ812に格納されている末尾位置情報812aが照合ターム識別情報として出力されることになる。よって、照合ターム識別情報は図12に示すように32ビットの照合ターム識別子と、16ビットの先頭位置情報および16ビットの末尾位置情報からなる32ビットの照合位置情報として出力されることになる。

【0041】以上の説明より明らかなように、照合情報は（Tn, Xs, Xe）と表すことができる。ここでTnは文書識別子または照合ターム識別子を表す。また、Xsは照合タームの先頭位置情報を、Xeは照合タームの末尾位置情報を表す。したがって図4の照合情報の例（T2, 33, 34）では、T2が照合ターム識別子を表し、33が照合タームの先頭位置情報を、34が照合タームの末尾位置情報を表すことになる。

【0042】以上説明したタームコンパレータ210、文書識別子検出回路220、文字数カウント回路230、検索ターム長テーブル250、および位置情報付加回路800の動作により文字列照合回路200からは図

13に示すような照合情報が文書毎に出力されることになる。すなわち、文書毎にまず先頭に文書識別情報を経て、次に照合ターム識別情報が並ぶ構成となる。また文書識別情報は、照合ターム識別情報と構造が同じであり、位置情報が0の照合ターム識別情報と見なすことができる。したがって、照合ターム識別情報と同様に扱うことができるため、文書識別情報を構造上意識することなく一括した処理を行うことが可能となる。以上が文字列照合回路200の詳細な説明である。

【0043】次に、文字列照合回路200の動作を具体例で説明する。ここでは、検索条件式

“Q=((文書[4C]理解)[S]システム)  
[AND](文書[S]検索)” (7-1)

を例として説明する。本例では検索タームとして“T1:文書”、“T2:理解”、“T3:検索”、“T4:システム”および“S1:。”の5つが検索制御手段101より送られ、検索タームはタームコンパレータ210に、検索タームの長さは検索ターム長テーブルに設定される。ここで、T1、T2、T3、T4、およびS1はそれぞれ検索ターム“文書”、“理解”、“検索”、“システム”、および“。”の照合ターム識別子を表わす。“S1:。”は、文脈条件の“[S]”すなわち文脈として文(センテンス)の指定に対応してセンテンスの文脈識別文字列(以後、文脈マーカと呼ぶ)としての“。”を検出するためのものである。

【0044】文書としては、

“...。文書理解を用いた検索システムである。...”

(7-2)

が入力されるものとする。文書識別子はD1とする。この文書データが文字列照合回路200へ入力されたときの照合結果205が図14に示すような以下の照合情報が出力される。

(D1, 0, 0) (8-1)

(S1, 30, 30) (8-2)

(T1, 31, 32) (8-3)

(T2, 33, 34) (8-4)

(T3, 39, 40) (8-5)

(T4, 41, 44) (8-6)

(S1, 48, 48) (8-7)

(8-1)は文書識別情報を表している。文書識別情報(8-1)においてD1は文書識別子を表し、これに続く2項は定数0である。(8-2)と(8-7)は文脈マーカ“。”の、(8-3)は“文書”の、(8-4)は“理解”の、(8-5)は“検索”の、および(8-6)は“システム”の照合ターム識別情報を表している。また、S1は“。”の、T1は“文書”の、T2は“理解”の、T3は“検索”の、およびT4は“システム”の照合ターム識別子を表している。これら(8-1)～(8-7)の照合情報205が複合判定回路3

00の入力として送られることになる。

【0045】次に、複合条件判定回路300の条件判定処理について説明する。複合条件判定回路300は図1に示すように、3つのマイクロコンピュータのMPUa301、MPUb302、およびMPUc303から構成されている。マイクロコンピュータMPUa301では近傍条件判定プログラム310が、マイクロコンピュータMPUb302では文脈条件判定プログラム320が、マイクロコンピュータMPUc303では論理条件判定プログラム330が実行される。さらに各MPU間にはファーストイン・ファーストアウト(FIFO)メモリを使用してバッファ350、360、および370が配置され、それぞれのMPU間のデータの受渡しに用いられている。

【0046】まず、近傍条件判定プログラム310の判定処理について説明する。近傍条件判定プログラム310では、文字列照合回路200からバッファ350に送り込まれた照合情報205を読み出し、検索情報202として指定された近傍条件に合致するか否かを判定する。近傍条件例としては、(7-1)の中に“文書[4C]理解”という条件式がある。“文書[4C]理解”という条件式は、“文書”と“理解”という2つの検索タームがこの順序で現われ、且つこの2つの検索タームが4文字以内に近接して現れる文書を探し出すということを表す。ここでは検索式を一種の演算式と見なして、“文書”を前方オペランドTa、“検索”を後方オペランドTbと、また“[4C]”をオペレーションと呼ぶことにする。さらに本近傍条件を表す識別子を“Pi”とする。なおPiには照合ターム識別子とは区別できるコードを割り付ける。このように定義することにより、近傍条件は“Pi:Ta[nC]Tb”と記述することができる。以下の説明はこの定義を用いて行う。

【0047】近傍条件のオペレーションにはこの他にも前述したように、

“Pi:Ta[nC,mC]Tb”

“Pi:Ta<nC>Tb”

“Pi:Ta[nc]Tb”

“Pi:Ta[nc,mc]Tb”

“Pi:Ta<nc>Tb”

などがある。この近傍条件の処理の手順について図15を用いて詳細に説明する。まず、繰返し処理ステップ1000ではバッファ350内の照合情報205を全て、すなわち最後の文書の照合情報を読み出し尽くすまで1001から1010までの処理ステップを繰り返す。

【0048】照合情報読み込み処理ステップ1001では、バッファ350より照合情報205を1個読み出し、バッファ360へ出力する。照合情報識別処理ステップ1002では、上記照合情報読み込み処理ステップ1002で読み込んだ照合情報205が文書識別情報なのか、あるいは照合ターム識別情報なのか調べる。すなわ

ち照合情報の下位32ビットが0(ゼロ)の場合には、文書識別子と判定する。文書識別子と判定した場合には、文書単位の初期化処理ステップ1004を実行する。ここではワークエリアとして使用する前方オペランドバッファ311の0(ゼロ)クリアを行う。照合情報205が照合ターム識別情報の場合には、後方オペランド識別処理ステップ1003を実行する。後方オペランド識別処理ステップ1003では、照合ターム識別情報の照合ターム識別子を調べて、後方オペランドとして近傍条件に指定された照合タームか否かを判定する。後方オペランドの場合には後述する前方オペランドバッファ311に格納されている照合ターム識別情報との距離を求め、指定された近傍条件を満たしているか否かを判定する。

【0049】以降、前方オペランドに指定された検索タームに関する照合ターム識別情報を前方照合ターム識別情報と呼び、後方オペランドに指定された検索タームに関する照合ターム識別情報を後方照合ターム識別情報と呼ぶことにする。すなわち、前方オペランドバッファ311に格納されている前方照合ターム識別情報と後方オペランドに指定されている後方照合ターム識別情報との近接条件を判定することになる。この近接条件判定では、まず前方照合ターム識別情報読み込み処理ステップ1006において前方オペランドバッファ311から前方照合ターム識別情報を読み込む。次に近接条件判定処理ステップ1007で、読み込まれた前方照合ターム識別情報と後方照合ターム識別情報の位置情報を比較して指定された近接条件が満足されるか否かを判定し、近接条件が満足される場合には判定結果を照合情報としてバッファ360および前方オペランドバッファ311に出力する。

【0050】前方オペランド識別処理ステップ1009では、照合ターム識別情報の照合ターム識別子が前方オペランドとして指定されている場合には、後方オペランドに指定されているものも含めて、この照合ターム識別情報を前方オペランドバッファ311に出力する。これは例えば、近傍条件として“P i : T a [n C] T b”と“P j : T b [n C] T c”が指定された場合、つまりT bが後方オペランドとしても、前方オペランドとしても指定されているような場合に、T bについては後方オペランド識別処理ステップ1003と前方オペランド識別処理ステップ1009の両方の処理が必要となるからである。このために、後方オペランド識別処理ステップ1003と前方オペランド識別処理ステップ1009を分けて処理を行うことになる。以上の各処理ステップをバッファ350に格納されている照合情報に対し繰返し実行していくことにより、近傍条件判定処理を実現することができる。

【0051】以上の処理手順を具体例で説明する。検索

条件としては、(7-1)に示した式“Q=((文書[4C]理解)[S]システム)[AND](文書[S]検索)”を例に用いて説明する。各条件判定プログラムには、検索制御手段101で解析され各条件に分離された条件式が設定される。具体的には、近傍条件判定プログラム310に本条件式(7-1)の近傍条件部分“文書[4C]理解”が、近傍条件識別子P1と“文書”および“理解”に対応する検索ターム識別子T1およびT2を用いて、“P1:T1[4C]T2”という形で与えられる。今、(7-2)に示した文書

“...。文書理解を用いた検索システムである。...”

が入力されたとすると、前述したように文字列照合回路200からは以下の(8-1)~(8-7)が照合情報としてバッファ350へ出力される。これらの照合情報を図17に示す。

【0052】(D1, 0, 0) (8-1)  
(S1, 30, 30) (8-2)  
(T1, 31, 32) (8-3)  
(T2, 33, 34) (8-4)  
(T3, 39, 40) (8-5)  
(T4, 41, 44) (8-6)  
(S1, 48, 48) (8-7)

(8-1)は文書識別情報を表している。D1は文書識別子である。(8-2)と(8-7)は“。”の照合ターム識別情報を表している。また、(8-3)、(8-4)、(8-5)、および(8-6)は、それぞれ“文書”、“理解”、“検索”、および“システム”の照合ターム識別情報を表している。さらに、S1, T1, T2, T3, およびT4は、それぞれ“。”、“文書”、“理解”、“検索”、および“システム”の照合ターム識別子を表している。以上の条件における近傍条件判定処理を図15を用いて1ステップずつ説明する。初期状態では図16a、図16bの初期状態に示すようにバッファ350に照合情報(8-1)~(8-7)が格納されており、前方オペランドバッファ311及びバッファ360は0クリアされた状態になっている。

【0053】近傍条件判定処理プログラムではこれらの照合情報205をバッファ350から一つずつ読み、近傍条件“P1:T1[4C]T2”について判定処理を行う。まずステップ1として、読み込み処理ステップ1001が実行され、照合情報(8-1)すなわち(D1, 0, 0)が図16のステップ1に示すようにプログラムのワークエリアに読み込まれ、バッファ360へ照合情報としてそのまま出力される。次に照合情報識別処理ステップ1002が実行され、照合情報(D1, 0, 0)が文書識別情報かどうか調べられる。照合情報(D1, 0, 0)は後部の2項が両方とも0(ゼロ)、すなわち下位32ビットが0(ゼロ)のため、文書識別情報と判断される。したがって、初期化処理ステップ100

4が実行され、内部のワークエリアである前方オペランドバッファ311がゼロクリアされる。その後ステップ2として再び、照合情報読み込み処理ステップ1001が実行され照合情報(8-2)すなわち(S1, 30, 30)が読み込まれ、同様にしてバッファ360へそのまま出力される。次に照合情報識別処理ステップ1002が実行され、照合情報(8-2)が文書識別情報か照合ターム識別情報かどうか調べられる。照合情報(8-2), すなわち(S1, 30, 30)は下位32ビットが0(ゼロ)でないため、文書識別情報ではなく照合ターム識別情報と判断される。そして、次の後方オペランド識別処理ステップ1003で、本照合ターム識別情報が近傍条件中に指定された後方オペランドに指定されたものに該当するかどうか調べられる。

【0054】近傍条件“P1:T1[4C]T2”の後方オペランドに指定された検索タームはT2であり、本照合タームS1は後方オペランドに該当しないため、1005~1008の処理は行われず、すなわち、近傍条件判定処理は行われず、次の前方オペランド識別処理ステップ1009が実行されることになる。本処理ステップでは、上記照合ターム識別情報(8-2)が近傍条件“P1:T1[4C]T2”の前方オペランドとして指定されたものに該当するかどうか調べられる。本照合タームはS1のため前方オペランドには該当しないため、処理ステップ1010は実行されず、すなわち前方オペランドバッファ311へ格納されることなく処理を終える。

【0055】ステップ3として繰返し処理1000により、読み込み処理ステップ1001が実行され、3番目の照合情報(8-3)すなわち(T1, 31, 32)が読み込まれ、第2の入力と同様にバッファ360へ出力される。同時に照合情報識別処理ステップ1002が実行される。文書識別情報か照合ターム識別情報か調べられる。(8-3)は照合ターム識別情報なので後方オペランド識別処理ステップ1003が実行され、後方オペランドでないため前方オペランド識別処理ステップ1009が実行される。照合ターム識別情報(8-3), すなわち(T1, 31, 32)は近傍条件“P1:T1[4C]T2”の前方オペランドに指定されているので、前方オペランド格納処理ステップ1010が実行され、図16のステップ3に示すように前方オペランドバッファに格納される。

【0056】また、ステップ4として繰返し処理1000により、読み込み処理ステップ1001が実行され、4番目の照合情報(8-4)すなわち(T2, 33, 34)が読み込まれ、第3の入力と同様にバッファ360へ照合ターム識別情報(8-4)が出力される。次に照合情報識別処理ステップ1002が実行され、文書識別情報か照合ターム識別情報か調べられる。(8-4)は照合ターム識別情報なので後方オペランド識別処理ス

テップ1003が実行される。(8-4)すなわち(T2, 33, 34)は近傍条件“P1:T1[4C]T2”の後方オペランドに指定されているので、処理ステップ1005から1008までの近接条件判定処理が実行される。まず前方オペランドバッファ読み込み処理ステップ1006が実行され、前方オペランドバッファ311に格納されている前方照合ターム識別情報(8-3)が読み込まれる。次に近接条件判定処理ステップ1007が実行され前方照合ターム識別情報(8-3)と後方照合ターム識別情報(8-4)との文字距離が調べられる。(8-3)すなわち(T1, 31, 32)の末尾位置Xeは32であり、(8-4)すなわち(T2, 33, 34)の先頭位置は33なので、その間の文字距離は0であり指定条件の4文字以下を満たしている。このため判定結果出力処理ステップ1008が実行され、判定結果として(8-3)の先頭位置である31を先頭位置とし、(8-4)の末尾位置である34を末尾位置とし、照合ターム識別子をP1とする近傍条件識別情報(P1, 31, 34)(図17)が図16a、図16bのステップ4のように前方オペランドバッファ311とバッファ360へ出力される。さらに、繰返し処理1000により、読み込み処理ステップ1001が実行され、5番目の照合情報(8-5)すなわち(T3, 39, 40)が読み込まれ、第4の入力と同様にバッファ360へ(8-5)が出力される。次に照合情報識別処理ステップ1002が実行され、文書識別情報か照合ターム識別情報か調べられる。照合ターム識別情報(8-5)は近傍条件に設定されていない照合識別情報であるため、後方オペランド識別処理ステップ1003および前方オペランド識別処理ステップ1009のいずれの処理も行われずに次の入力に移る。以後、最後の照合ターム情報(8-7)まで同じように近傍条件判定処理を繰り返す。

【0057】以上の近傍条件判定処理によって、以下に示す(9-1)~(9-8)の照合ターム識別情報(図17)がバッファ360へ出力される。

(D1, 0, 0) (9-1)  
(S1, 30, 30) (9-2)  
(T1, 31, 32) (9-3)  
(T2, 33, 34) (9-4)  
(P1, 31, 34) (9-5)  
(T3, 39, 40) (9-6)  
(T4, 41, 44) (9-7)  
(S1, 48, 48) (9-8)

ここで注目すべき点は、近傍条件判定結果(9-5)も照合ターム識別情報として、末尾位置情報の昇順にソートし、格納されている点である。このため、後述するように文脈条件判定の際、全ての文脈と照合タームとの組合せについて包含関係をチェックしなくても済むため処理を軽減できるようになる。これらの照合ターム識別情

報は文脈条件判定プログラム320へ送られることになる。

【0058】次に、文脈条件判定プログラム320の判定処理について説明する。文脈条件判定プログラム320では、近傍条件判定プログラム310からバッファ360に送り込まれた照合情報を読み出し、検索情報202として指定された文脈条件に合致するか否かを判定する。文脈条件の例としては、(7-1)に示したような“文書[S] 検索”がある。“文書[S] 検索”という条件式は、“文書”と“検索”という2つの検索タームがこの順序で、同一の文(センテンス)に現れる文書を探し出すということを表す。ここで本文脈条件を表す識別子(以後、文脈条件識別子と呼ぶ)をCiとする。Ciには照合ターム識別子とは区別できるコードを割り付ける。このように定義することにより、文脈条件は“Ci:Ta[S]Tb”と記述することができる。以下の説明はこの定義を用いて行うことにする。文脈条件にはこの他にも、

“Ci:Ta[P]Tb”,  
“Ci:Ta[PH]Tb”,  
“Ci:Ta[p]Tb”,  
“Ci:Ta[s]Tb”,  
“Ci:Ta[ph]Tb”

などがある。

【0059】文脈条件判定の原理について、図18の近傍条件判定処理出力例を用いて説明する。(7-2)に示した文書“... 文書理解を用いた検索システムである。...”が入力されたことを想定すると、前述したように近傍条件判定プログラム310からは本図に示す以下の(9-1)～(9-8)が照合情報としてバッファ360を介し文脈条件判定処理プログラム320へ送られる。

(D1, 0, 0) (9-1)  
(S1, 30, 30) (9-2)  
(T1, 31, 32) (9-3)  
(T2, 33, 34) (9-4)  
(P1, 31, 34) (9-5)  
(T3, 39, 40) (9-6)  
(T4, 41, 44) (9-7)  
(S1, 48, 48) (9-8)

(9-1)は文書識別情報を表している。D1は文書識別子を表し、これに続く2項は定数0である。(9-2)と(9-8)は文脈マーカー“.”の照合ターム識別情報を表している。同様に、(9-3)、(9-4)、(9-6)、および(9-7)は、それぞれ“文書”、“理解”、“検索”“システム”の照合ターム識別情報を表している。ここで、S1, T1, T2, T3, およびT4は、それぞれ“.”、“文書”、“理解”、“検索”、および“システム”の照合ターム識別子を表す。また、(9-5)は近傍条件“文書[4C]

理解”が照合された際の近傍条件識別情報を表している。P1は近傍条件“文書[4C] 理解”の近傍条件識別子である。

【0060】以上の照合情報に基づいて、まず照合ターム識別子の順序が指定されている文脈条件“Ci:Ta[S]Tb”に関する共起判定を実行する。ここで識別子Taを0番目の識別子と呼び、識別子Tbを1番目の識別子と呼ぶ。共起判定では、文脈条件に指定されている全ての識別子が文脈中に出現したかどうかの判定を行う。共起判定の成否を判断するために、ここでは共起カウンタを用いる。この共起カウンタは指定された文脈条件に対応して1つずつ設ける。これより順序が指定されている文脈条件における共起カウンタの制御方法について、以下説明する。文脈条件“Ci:Ta[S]Tb”における文脈すなわち文(センテンス)の範囲は、出現した文脈マーカー“.”の照合情報(9-2)から次に出現する文脈マーカー“.”の照合情報(9-8)までである。よって、(9-2)～(9-8)、すなわち(S1, 30, 30)～(S1, 48, 48)がセンテンスの範囲となる。このため、(9-2)すなわち(S1, 30, 30)から順番に調べる。文脈マーカー“.”が出現すると共起カウンタを0(ゼロ)にリセットし、次に文脈条件に指定された照合情報が現れるかどうかを調べていく。まず、文脈条件に指定された第0番目の識別子Taが現れるまで共起カウンタの値を変更しない。すなわち、共起カウンタの値が0のとき第0番目の識別子Taに着目し、この識別子が現れるのを監視する。ここで識別子Taの照合情報が現れた場合、共起カウンタをカウントアップする。つまり共起カウンタの値を0から1に変更する。

【0061】次に共起カウンタの値が1を示しているので、文脈条件に指定されている第1番目の識別子Tbが現れるのを監視し、現れた場合に共起カウンタをカウントアップする。ここで識別子Tbの照合情報が現れた場合、共起カウンタをカウントアップする。すなわち、共起カウンタの値を1から2に更新する。ここで共起カウンタの値が2になったところで、本文脈条件に指定された識別子が全て現れたことになるため、共起条件が成立したことを判別できる。このように共起カウンタを制御することにより、順序が指定された文脈条件における共起判定を行っている。共起カウンタの判定は、後述するように次の文脈マーカーが現れた時点で行う。さらに、次に文脈マーカー“.”の照合情報(9-8)すなわち(S1, 48, 48)が出現する。この時点の共起カウンタの値は2になっている。共起カウンタの値が2ということは、文脈条件に指定されている2つの検索タームの識別子が出現し、共起条件が成立していることを表している。このとき、成立した文脈条件の文脈条件識別情報を出力する。この文脈条件識別情報の先頭位置情報には前に現れた文脈マーカー“.”の先頭位置情報で



ある30を設定し、末尾位置情報には後に現れた文脈マーカ―“.”の末尾位置情報である48を設定する。また、文脈条件識別情報の識別子には文脈条件の識別子“C1”を設定する。すなわち、文脈条件識別情報(C1, 30, 48)を出力する。

【0062】次に、照合ターム識別子の順序を問わない文脈条件“Cj:Ta[s]Tb”の場合の共起カウンタの制御方法について、以下説明する。文脈の範囲は同様に(9-2)~(9-8)、すなわち(S1, 30, 30)~(S1, 48, 48)がセンテンスの範囲となる。このため、同様に(9-2)すなわち(S1, 30, 30)から順番に調べる。まず、文脈マーカ―“.”が出現すると同様に共起カウンタを0(ゼロ)にリセットする。次に文脈条件に指定された照合情報が現れるかどうかを調べていく。まず、識別子“Tb”の照合情報が現れた場合、共起カウンタをカウントアップする。つまり共起カウンタの値は0から1に更新する。この後で既に出現した識別子“Tb”の照合情報が同一文脈中に再び現れた場合には共起カウンタの値を変更しない。次に識別子“Ta”の照合情報が現れた場合、共起カウンタをカウントアップする。つまり共起カウンタの値は1から2に更新する。ここで共起カウンタの値が2になったところで、本文脈条件に指定された2つの識別子が全て現れたことになるため、共起条件が成立したことが判別できる。この後で既に出現した識別子“Ta”の照合情報が同一文脈中に再び現れた場合にも共起カウンタの値を変更しない。さらに、次に文脈マーカ―“.”の照合情報(9-8)すなわち(S1, 48, 48)が出現したとき、既に共起カウンタが2、すなわち共起条件が成立しているため、本文脈条件の文脈条件識別情報を出力する。すなわち、文脈条件識別情報(Cj, 30, 48)を出力する。このように共起カウンタを制御することにより、順序を問わない文脈条件の共起判定にも使用することができる。以上のように共起カウンタを制御することにより共起判定の成否を判断することができる。

【0063】文脈条件の処理の手順について図19を用いて詳細に説明する。まず繰返し処理ステップ1100ではバッファ360内の照合情報を全て、すなわち最後の文書の照合情報を読み出し尽くすまで1101から1112までの処理ステップを繰り返す。照合情報読み込み処理ステップ1101では、バッファ360より照合情報を1個読み出し、ワークエリアへ出力する。照合情報識別処理ステップ1102では、上記照合情報読み込み処理ステップ1101で読み込んだ照合情報が照合ターム識別情報なのかどうかを調べる。すなわち照合情報の下位32ビットが0(ゼロ)でない場合には、照合ターム識別子と判定する。この場合には、文脈マーカ―を検出する文脈マーカ―識別処理ステップ1103を実行する。照合ターム識別情報以外の場合には、照合情報をバ

ッファ370へ出力する照合情報出力処理ステップ1112を実行する。文脈マーカ―識別処理ステップ1103では、照合ターム識別情報の照合ターム識別子を調べて文脈条件に指定された文脈の文脈マーカ―が否かを判定する。文脈マーカ―の場合には、後処理ステップ1104が実行される。

【0064】後処理ステップ1104では、これまで共起判定を行ってきた文脈の末尾位置情報を格納し、ワークエリアに格納する。この文脈識別情報は後述する照合ターム識別情報出力処理ステップ1110でバッファ370に出力される。その後、前処理ステップ1104aを実行する。ここでは本文脈マーカ―を端点とする次の文脈に対し共起判定を実行するための準備を行う。まず、本文脈マーカ―を端点とする文脈に関する文脈条件の共起カウンタを0にリセットする。さらに、本文脈マーカ―の先頭位置情報を、本文脈マーカ―を端点とする文脈の先頭位置情報とする。後処理ステップ1104において、文脈条件識別情報をワークエリアに格納し、再度文脈条件判定を行うのは文脈条件が入れ子の場合を考慮しているからである。文脈条件の入れ子とは、すなわち、“Cj:(Ta[s]Tb)[p]Tc”のように“Ta”と“Tb”という2つの検索タームがその順序を問わず、同一文(センテンス)に現れ、且つこの文(センテンス)と“Tc”という検索タームが順序を問わず、同一段落(パラグラフ)に現れる文書を探し出すというように文脈条件中に文脈条件が指定される場合をいう。

【0065】この判定手順としては、文脈条件を“Ci:(Ta[s]Tb)”と“Cj:Ci[p]Tc”とに分け、まず含まれている方の文脈条件すなわち本例では段落に包含されているセンテンスに関する文脈条件“Ci:(Ta[s]Tb)”を判定する。このセンテンスの文脈条件が成立したと想定する。この成立時にセンテンスの文脈条件識別情報が一時格納バッファに格納される。次にセンテンスの文脈マーカ―が現れたとき、このセンテンスの文脈条件識別情報に末尾位置情報が設定され、文脈条件識別情報として確定する。このセンテンスの文脈条件識別情報をワークエリアに格納し、含む方の文脈条件すなわち本例では段落に関する文脈条件“Cj:Ci[p]Tc”の判定対象とする。その後、照合ターム識別子Tcの照合ターム識別情報が現れたと想定する。このとき、段落の文脈条件は成立し、一時格納バッファに段落の文脈条件識別情報が格納される。次に段落の文脈マーカ―が出現したとき、この段落の文脈条件識別情報に末尾位置情報が設定され、文脈条件識別情報として確定される。このように処理を行うことによって、入れ子の文脈条件“Cj:(Ta[s]Tb)[p]Tc”の条件判定を実現している。

【0066】文脈マーカ―識別処理ステップ1103の後には、ワークエリア繰返し処理ステップ1106を実

29

行する。ワークエリア繰返し処理ステップ1106では、照合情報読み込み処理ステップ1101や後処理ステップ1104でワークエリアに格納された全ての照合ターム識別情報について1107~1110の共起判定処理を繰返し実行する。文脈条件指定識別処理ステップ1107では、ワークエリアに格納されている照合ターム識別情報の照合ターム識別子から文脈条件に指定されているか否かを調べる。文脈条件に指定されている場合には共起判定処理ステップ1108が実行され、文脈条件に指定されている照合ターム識別子が文脈中に現れる度にカウントアップする共起カウンタ値が2かどうかを調べる。共起カウンタ値が2ということは、文脈条件に指定された2つの照合ターム識別子が発見され文脈条件が成立したことを表している。文脈条件が成立した場合には、該当文脈条件の文脈条件識別子を一時格納バッファ321に格納する一時格納処理ステップ1109を実行する。その後、文脈マーカ識別処理ステップ1110aを実行し、文脈マーカの照合ターム識別情報でなければ照合ターム識別情報出力処理ステップ1110を実行する。ここでは、照合ターム識別情報をバッファ370に出力する。以上の各処理ステップをバッファ360に格納されている照合情報に対し繰返し実行していくことにより文脈条件判定処理を実現することができる。

【0067】以上の処理手順を具体例で説明する。検索条件としては、(7-1)に示した式“Q=((文書[4C]理解)[S]システム)[AND](文書[s]検索)”を例に用いる。各条件判定プログラムには、検索制御手段101で解析され各条件に分離された条件式が設定される。具体的には、文脈条件判定プログラム320には本条件式(7-1)の文脈条件部分“(文書[4C]理解)[S]システム”と“文書[s]検索”が設定される。ここでは、近傍条件“文書[4C]理解”の識別子をP1とし、“システム”の検索ターム識別子をT3とすることにより、文脈条件“(文書[4C]理解)[S]システム”を“C1:P1[S]T3”という形で表す。C1は文脈条件の識別子である。同様に、文脈条件“文書[s]検索”は“C2:T1[s]T4”と表される。C2は文脈条件の識別子であり、T1は“文書”の、T4は“検索”の検索ターム識別子である。

【0068】今、(7-2)に示した文書  
“...文書理解を用いた検索システムである...”

が入力されたとすると、前述したように近傍条件判定プログラム310からは図18に示す以下の(9-1)~(9-8)が照合情報としてバッファ360へ出力される。

(D1, 0, 0) (9-1)  
(S1, 30, 30) (9-2)  
(T1, 31, 32) (9-3)

30

(T2, 33, 34) (9-4)  
(P1, 31, 34) (9-5)  
(T3, 39, 40) (9-6)  
(T4, 41, 44) (9-7)  
(S1, 48, 48) (9-8)

本例では、文脈マーカの照合情報(9-2)と(9-8)を端点とする文脈、すなわちセンテンスにおいて、文脈条件“(文書[4C]理解)[S]システム”と文脈条件“文書[s]検索”が成立するか否かが判定されることになる。

【0069】この条件式における文脈条件判定処理について図20a、図20bと図21a、図21bを用いて説明する。まず、図20a、図20bに示す初期状態においては、バッファ360に照合情報(9-1)~(9-8)が格納されており、一時格納バッファ321及びバッファ370は0クリアされた状態になっている。文脈条件判定処理プログラム320ではこれらの照合情報をバッファ360から照合情報をつつ読み込み、文脈条件“C1:P1[S]T3”と文脈条件“C2:T1[s]T4”について判定処理を行う。まずステップ1として、図19に示す読み込み処理ステップ1101が実行され、照合情報(9-1)すなわち照合情報(D1, 0, 0)が図20a、図20bのステップ1に示すようにプログラムのワークエリアに読み込まれる。次に照合情報識別処理ステップ1102が実行され、照合情報(9-1)が照合ターム識別情報かどうか調べられる。照合情報(D1, 0, 0)は後部の2項が両方とも0(ゼロ)、すなわち下位32ビットが0(ゼロ)のため、文書識別情報と判断される。したがって、照合情報(D1, 0, 0)は照合ターム識別情報でないため、文脈マーカ識別処理ステップ1110aにより文脈マーカのものとないと判断される。このため、次の照合情報出力処理ステップ1110が実行され、そのままバッファ370へ出力される。

【0070】その後ステップ2として、再び照合情報読み込み処理ステップ1101が実行され、照合情報(9-2)すなわち(S1, 30, 30)が読み込まれる。次に照合情報識別処理ステップ1102が実行され、照合情報(S1, 30, 30)が照合ターム識別情報かどうか調べられる。照合情報(S1, 30, 30)は下位32ビットが0(ゼロ)でないため、文書識別情報ではなく照合ターム識別情報と判断される。そして、次の文脈マーカ識別処理ステップ1103が実行され、本照合ターム識別情報が文脈条件中に指定された文脈の文脈マーカに該当するかどうか調べられる。照合情報(S1, 30, 30)の照合ターム識別子S1は文脈マーカに指定された“.”に対応するものであるため、後処理ステップ1104が実行されることになる。

【0071】後処理ステップ1104では、まず本文脈マーカS1を端点とする文脈の末尾位置情報を設定す

る。次に、一時格納バッファ321に照合情報が格納されているかどうかを調べ、格納されている場合には一時格納バッファ321の内容を全てワークエリアに読み込む。一時格納バッファ321には、後述するように指定文脈条件に対応した文脈条件識別情報が照合情報として格納されている。ここでは、照合情報が格納されていないため、ワークエリアには読み込まれないことになる。この後に、前処理ステップ1104aを実行する。ここではセンテンスの文脈条件、すなわち“C1:P1[S]T3”と“C2:T1[s]T4”に対応する共起カウンタaと共起カウンタbに0を設定する。次に文脈先頭位置情報に文脈マーカース1の先頭位置情報である30を設定する。その後、ワークエリア繰返し処理ステップ1106を実行する。ここでは本照合ターム識別子“S1”が文脈条件に指定されていないため文脈条件指定識別処理ステップ1107は実行されない。さらに文脈マーカース1の識別処理ステップ1110aが実行されるが、本照合ターム識別子“S1”は文脈マーカース1の照合ターム識別情報であるため、照合ターム識別情報出力処理ステップ1110は実行されない。したがって、文脈マーカース1の照合ターム識別情報はバッファ370に出力されないことになる。

【0072】ステップ3として、繰返し処理ステップ1100により読込み処理ステップ1101が実行され、3番目の照合情報(9-3)、すなわち(T1, 31, 32)がワークエリアに読み込まれる。次に照合情報識別処理ステップ1102が実行され、照合ターム識別情報か否かが調べられる。照合情報(T1, 31, 32)は照合ターム識別情報なので文脈マーカース1の識別処理ステップ1103が実行される。照合ターム識別情報(T1, 31, 32)の照合ターム識別子T1は文脈マーカース1の照合ターム識別子S1でないため、後処理ステップ1104は実行されない。この後、ワークエリア繰返し処理ステップ1106が実行され、ワークエリアに格納されている照合ターム識別情報について1107~1110の共起判定処理が実施される。文脈条件指定識別処理ステップ1107では、ワークエリアに格納されている照合ターム識別情報(T1, 31, 32)の照合ターム識別子T1を参照して文脈条件に指定された照合タームか否かを調べる。本照合ターム識別子T1は、文脈条件“C2:T1[s]T4”に指定されているので共起判定処理1108が実行されることになる。ここで、文脈条件“C2:T1[s]T4”に対応した共起カウンタaはカウントアップされ、共起カウンタaの値は0から1に更新される。しかし、共起カウンタaの値が2でないため、共起判定は成立しないことになる。この後、文脈マーカース1の識別処理ステップ1110aを実行するが文脈マーカース1の照合ターム識別情報でないため、照合ターム識別情報出力処理ステップ1110が実行される。ここで照合ターム識別情報(T1, 31, 32)がバッ

ファ370に出力される。

【0073】また、ステップ4として繰返し処理ステップ1100のもとに、読込み処理ステップ1101が実行され、4番目の照合情報(9-4)、すなわち(T2, 33, 34)がワークエリアに読み込まれる。そして照合情報識別処理ステップ1102が実行され、照合ターム識別情報か否かが調べられる。照合情報(T2, 33, 34)は照合ターム識別情報なので文脈マーカース1の識別処理ステップ1103が実行されるが、照合ターム識別情報(T2, 33, 34)の照合ターム識別子はS1でないため、後処理ステップ1104は実行されない。この後、ワークエリア繰返し処理ステップ1106のもとに、文脈条件指定識別処理ステップ1107が実行され、ワークエリア内の照合ターム識別情報が文脈条件に指定されているか否かが調べられる。ワークエリアに格納されている照合ターム識別情報(T2, 33, 34)の照合ターム識別子T2は文脈条件に指定されていないため、共起判定処理ステップ1108は実行されない。また、照合ターム識別情報は文脈マーカース1でないため、照合ターム識別情報出力処理ステップ1110が実行され、照合ターム識別情報(T2, 33, 34)がバッファ370に出力される。

【0074】さらにステップ5として、繰返し処理ステップ1100のもとで、読込み処理ステップ1101が実行され、5番目の照合情報(9-5)すなわち(P1, 31, 34)がワークエリアに読込まれる。更に照合情報識別処理ステップ1102により照合ターム識別情報か否かが調べられ、照合情報(P1, 31, 34)は照合ターム識別情報なので文脈マーカース1の識別処理ステップ1103が実行される。ここで、照合ターム識別情報(P1, 31, 34)の照合ターム識別子はS1でないため、この後ワークエリア繰返し処理ステップ1106が実行され、ワークエリアに格納されている照合ターム識別情報について1107~1110の共起判定処理が実行される。文脈条件指定識別処理ステップ1107では、ワークエリアに格納されている照合ターム識別情報(P1, 31, 34)の照合ターム識別子P1を参照して文脈条件に指定されているか否かが調べられる。この場合、文脈条件“C1:P1[S]T3”に対応する共起カウンタbの値が0であり、且つ照合ターム識別子P1は文脈条件“C1:P1[S]T3”の第0番目の照合ターム識別子として指定されている。このため共起判定処理1108が実行されることになる。ここで、共起カウンタbはカウントアップされ、共起カウンタbの値は0から1になる。しかし、共起カウンタbの値が2でないため、共起判定は成立しないことになる。この後、本照合ターム識別情報が文脈マーカース1のものでないため照合ターム識別情報出力処理ステップ1110が実行され、本照合ターム識別情報(P1, 31, 34)がバッファ370に出力される。

【0075】さらにステップ6として、繰返し処理ステップ1100のもとに、読み込み処理ステップ1101が実行され、6番目の照合情報(9-6)すなわち(T3, 39, 40)がワークエリアに読み込まれ、その後、照合情報識別処理ステップ1102が実行される。ここで、照合情報(T3, 39, 40)が照合ターム識別情報か否かが調べられ、照合情報(T3, 39, 40)は照合ターム識別情報なので文脈マーカ識別処理ステップ1103が実行されるが、照合ターム識別情報(T3, 39, 40)は文脈マーカでないため、後処理ステップ1104は実行されないことになる。この後、ワークエリア繰返し処理ステップ1106が実行され、ワークエリアに格納されている照合ターム識別情報について1107~1110の共起判定処理が実行されることになる。文脈条件指定識別処理ステップ1107では、ワークエリアに格納されている照合ターム識別情報(T3, 39, 40)の照合ターム識別子T3を参照して、本照合ターム識別子が文脈条件に指定されているか否かが調べられる。この場合、文脈条件“C1:P1[S]T3”に対応する共起カウンタbの値が1であり、且つ照合ターム識別子T3は文脈条件“C1:P1[S]T3”の1番目の識別子に指定されているので共起判定処理1108が実行される。ここで共起カウンタbがカウントアップされ、共起カウンタbの値は1から2に更新される。共起カウンタbの値が2になったことから、共起判定が成立したものと判断することができる。このとき、一時格納処理ステップ1109が実行され、文脈条件“C1:P1[S]T3”の文脈条件識別情報(C1, 30, Xe1)が一時格納バッファ321に格納される。ここでは、文脈の後方を端点とする文脈マーカの位置情報がまだ判明しないため、仮に文脈末尾位置情報をXe1として置く。この後に文脈マーカの照合ターム識別情報が現れたとき、この文脈末尾位置情報は後処理ステップ1104で設定される。その後、本照合ターム識別情報は文脈マーカのものでないため照合ターム識別情報出力処理ステップ1110が実行され、今成立した照合ターム識別情報(T3, 39, 40)がバッファ370に出力される。

【0076】ステップ7として、繰返し処理ステップ1100のもとで、読み込み処理ステップ1101が実行され、7番目の照合情報(9-7)すなわち(T4, 41, 46)がワークエリアに読み込まれる。その後、照合情報識別処理ステップ1102が実行され、本照合情報が照合ターム識別情報か否かが調べられる。照合情報(T4, 41, 46)は照合ターム識別情報なので文脈マーカ識別処理ステップ1103が実行され、照合ターム識別情報(T4, 41, 46)は文脈マーカでないため、後処理ステップ1104は実行されない。この後、ワークエリア繰返し処理ステップ1106が実行され、ワークエリアに格納されている照合ターム識別情報

について1107~1110の共起判定処理が実行されることになる。文脈条件指定識別処理ステップ1107では、ワークエリアに格納されている照合ターム識別情報(T4, 41, 46)の照合ターム識別子T4を参照して、本照合タームが文脈条件に指定されているか否かが調べられる。この場合、照合ターム識別子T4は文脈条件“C2:T1[s]T4”に指定されているので共起判定処理1108が実行される。ここで、文脈条件“C2:T1[s]T4”におけるの照合ターム識別子T4が現れたため、本文脈条件に対応する共起カウンタaがカウントアップされ、共起カウンタaの値は1から2になり、共起判定が成立することになる。したがって、一時格納処理ステップ1109が実行され、今成立した文脈条件“C2:T1[s]T4”の文脈条件識別情報(C2, 30, Xe2)が一時格納バッファ321に格納される。文脈末尾位置情報は決定されていないのでXe2を一時的に設定しておく。その後、本照合ターム識別情報が文脈マーカのものでないため照合ターム識別情報出力処理ステップ1110が実行され、照合ターム識別情報(T4, 41, 46)がバッファ370に出力される。

【0077】最後にステップ8として、再び照合情報読み込み処理ステップ1101が実行され、本照合情報(9-8)すなわち(S1, 48, 48)が読み込まれる。次に照合情報識別処理ステップ1102が実行され、照合情報(S1, 48, 48)が照合ターム識別情報であるかどうか調べられる。照合情報(S1, 48, 48)は照合ターム識別情報であるため、次の文脈マーカ識別処理ステップ1103が実行され、本照合ターム識別情報が文脈条件中に指定された文脈に関する文脈マーカに該当するかどうか調べられる。照合ターム識別子はS1であるため指定されたセンテンスの文脈マーカ“。”に該当することになり、後処理ステップ1104が実行される。後処理ステップ1104では、まず本文脈マーカの末尾位置情報である48を文脈末尾位置情報とする。次に、センテンスの文脈条件識別情報(C1, 30, Xe1)と(C2, 30, Xe2)が格納されている一時格納バッファ321にセンテンスの文脈末尾位置情報48を設定し、この処理結果は図21a、図21bのステップ8に示した(C1, 30, 48)と(C2, 30, 48)のようになる。この処理結果はワークエリアに格納される。さらに前処理ステップ1104aが実行され、センテンスの文脈条件の共起カウンタを0クリアし、本文脈マーカの先頭位置情報48を文脈先頭位置情報に設定する。

【0078】その後、ワークエリア繰返し処理ステップ1106が実行される。まず文脈条件識別情報(C1, 30, 48)について、文脈条件指定識別処理ステップ1107が実行されるが照合ターム識別子C1は文脈条件に指定されていないため共起判定処理1108は実行

されず、照合ターム識別情報出力処理ステップ1110が実行され、文脈条件識別情報(C1, 30, 48)がバッファ370に出力される。すなわち、本文脈条件識別情報の位置情報は、文脈マーカーの識別子(9-2)と(9-8)を端点とするセンテンスの先頭位置情報30と末尾位置情報48を位置情報としている。この後に再度、ワークエリア繰返し処理ステップ1106が実行され、文脈条件識別情報(C2, 30, 48)について、文脈条件指定識別処理ステップ1107が実行されるが照合ターム識別子C2も文脈条件に指定されていないため、照合ターム識別情報出力処理ステップ1110により、文脈条件識別情報(C2, 30, 48)がバッファ370に出力される。

【0079】以上の文脈条件判定処理が実行されることにより、図22に示す(10-1)～(10-8)の照合情報がバッファ370へ出力される。

(D1, 0, 0) (10-1)  
(T1, 31, 32) (10-2)  
(T2, 33, 34) (10-3)  
(P1, 31, 34) (10-4)  
(T3, 39, 40) (10-5)  
(T4, 41, 44) (10-6)  
(C1, 30, 48) (10-7)  
(C2, 30, 48) (10-8)

ここで、(10-7)は文脈条件“(文書[4C]理解)[S]システム”の文脈条件識別情報であり、同様に(10-8)は“文書[s]検索”の文脈条件識別情報である。これらの照合情報は引き続き論理条件判定プログラム330へ送られることになる。

【0080】最後に、論理条件判定プログラム330の判定処理について説明する。論理条件判定プログラム330では、文脈条件判定プログラム320からバッファ370に送り込まれた照合情報を読み出し、検索制御手段101より検索情報202として指定された論理条件に合致するか否かを判定する。論理条件例としては、例えば“文書[AND]検索”という条件式がある。“文書[AND]検索”という条件式は、“文書”と“検索”という2つの検索タームが同時に現れる文書を探し出すということを表す。また、本論理条件を表す識別子(以後、論理条件識別子と呼ぶ)をLiとする。なおLiには照合ターム識別子とは区別できるコードを割り付ける。このように定義することにより、論理条件は“Li:Ta[AND]Tb”

と記述することができる。以下の説明はこの定義を用いて行う。論理条件にはこの他にも、“Li:Ta[OR]Tb”と“Li:Ta[NOT]Tb”がある。ここで論理条件“Li:Ta[OR]Tb”は、“Ta”あるいは“Tb”という検索タームが現れる文書を探し出すことを表し、論理条件“Li:Ta[NOT]Tb”は、“Ta”という検索タームが現われて、

且つ“Tb”という検索タームが現われない文書を探し出すことを表している。

【0081】これらの論理条件は検索制御手段101において、以下に示す積の一般形式に変換され、検索情報202として論理条件判定プログラム330に指定される。

Li:(A11+A12+...+A1j)  
\*(A21+A22+...+A2k)  
\*  
\*  
10 \*(An1+A22+...+Anm)  
(10-1)

(10-1)式において、“+”は論理和を表わし、“\*”は論理積を表わす。また、Anmを要素と呼び、(An1+A22+...+Anm)を項と呼ぶ。要素Anmとしては否定(“-”で表す)が掛かっている要素¬Anmも存在する。また、項にも否定が掛かった項¬(An1+A22+...+Anm)が存在する。ここで¬Anmを負論理の要素と呼ぶ。これに対し、否定が掛かっていない項を正論理の要素と呼ぶ。また、¬(An1+A22+...+Anm)を負論理の項と呼び、否定が掛かっていない項を正論理の項と呼ぶ。

【0082】(10-1)式では、項の論理積の形になっているので、(10-1)式が成立する(真になる)ためには、各々の項が全て成立する必要がある。そのため、カウンタを用意し、成立した項をカウントする。このカウンタ(以後、項カウンタと呼ぶ)の値が項の個数に等しければ、(10-1)式は成立したことになる。負論理の要素を含まない負論理の項と負論理の要素を含む正論理の項は最初から成立している。したがって、項カウンタの初期値には負論理の要素を含まない負論理の項の個数と負論理の要素を含む正論理の項の個数とを加算した値を設定する。この項カウンタを次のように制御することにより、(10-1)式の成否を判定することができる。つまり、項カウンタには初期値として負論理の項の個数を設定し、項が不成立から成立に変化した場合には項カウンタに1を加算し、逆に項が成立から不成立に変化した場合には項カウンタから1を減算する。項が不成立から不成立または成立から成立のように変化しない場合は項カウンタの値を更新しない。このように項カウンタを制御し、文書単位に項カウンタの値が項の個数に等しいかを調べることにより(10-1)式の成立を判別することができる。また、式全体に否定が掛っている場合には逆に、項カウンタの値が項の個数未満かを調べることにより(10-1)式の成立を判別することができる。

【0083】次に、全ての項では要素の論理和になっているため、項が成立するためにはいずれかの要素が成立していれば良いことになる。1つの項の成否判定方法について以下説明する。ここでは、項の成否を調べるためにカウンタを用い、項に指定されている要素において、

要素が成立したものの個数をカウントする。このカウンタ（以後、要素カウンタと呼ぶ）の値が1以上、すなわちいずれかの要素が成立している場合は、項は成立したものと見做す。負論理の要素は最初から成立していると思ふので、要素カウンタの初期値には負論理の要素の個数を設定する。このため、要素カウンタは項に対応して1つずつ、すなわち項の個数だけ設ける。

【0084】この要素カウンタを以下のように制御することにより項の成否を判定することができる。つまり、要素カウンタには初期値として負論理の要素の個数を設定し、要素が不成立から成立に変化した場合には要素カウンタに1を加算し、逆に要素が成立から不成立に変化した場合には要素カウンタから1を減算する。ここでの要素は照合ターム識別子に対応しているため、要素が不成立から不成立または成立から成立のように変化しない。よって、不成立から成立に変化するの正論理の要素の場合のみであり、成立から不成立に変化するの負論理の要素の場合のみである。したがって、正論理の要素に対応する照合ターム識別情報が現れた場合には要素カウンタに1を加算し、逆に負論理の要素に対応する照合ターム識別情報が現れた場合には要素カウンタから1を減算する。

【0085】このように要素カウンタを制御し、照合ターム識別情報が入力される度に要素カウンタの値が1以上かを調べることににより項の成立を判断する。また、項に否定が掛かっているときには逆に、要素カウンタの値が0かを調べることににより項の成立を判断する。この後、項の成立判定の結果をもとに、項が不成立から成立に変化した場合には項カウンタに1を加算し、逆に項が成立から不成立に変化した場合には項カウンタから1を減算する。また、否定を含めた項が不成立から不成立または成立から成立のように変化しない場合は項カウンタの値を更新しない。このように項カウンタを制御し、文書単位に項カウンタの値が項の個数に等しいか否かを調べることににより（10-1）式の成否を判別することができる。例えば上記の“Li:Ta[AND]Tb”すなわち“Li:Ta\*Tb”の場合には、

“Li:¬((¬Ta)+(¬Tb))”

のように変換され検索情報202として論理条件判定プログラム330に渡される。

【0086】この論理条件の処理の手順について図23を用いて詳細に説明する。例えば上記の“Li:Ta[AND]Tb”すなわち“Li:Ta\*Tb”の場合には、

“Li:¬((¬Ta)+(¬Tb))”

(10-2)

のように変換され検索情報202として論理条件判定プログラム330に渡される。本論理条件には、負論理の項が1つ存在する。さらにこの項の中には、負論理の要素が2つ存在する。項が1つなのでここでは、要素カウ

ンタを1つ使用する。項カウンタの初期値には、負論理の要素を含まない負論理の項の個数と負論理の要素を含む正論理の項の個数とを加算した値を設定することになるが、本例の項は負論理要素を含む負論理の項なので該当しないため0を設定する。また、要素カウンタには、負論理の要素の個数である2を設定する。ここでは、項に¬が掛かっているため、要素カウンタをカウントダウンし、要素カウンタの値が0になったときに項が成立したものと判断することができる。また、論理条件（10-2）式に¬が掛かっていないため、項カウンタの値が1のときに論理条件（10-2）式が成立したものと判断することができる。

【0087】まず繰返し処理ステップ1200ではバッファ370内の照合情報を全て、すなわち最後の文書の照合情報を読み出し尽くすまで1201から1210までの処理ステップを繰り返す。読み込み処理ステップ1201では、バッファ370より照合情報を読み出しプログラムのワークエリアへ格納する。照合情報識別処理ステップ1202では、上記読み込み処理ステップ1201で読み込んだ照合情報が文書識別情報か、それとも照合ターム識別情報なのかを調べる。すなわち照合情報の後部2項が両方とも0（ゼロ）、すなわち下位32ビットが0（ゼロ）の場合には、文書識別情報と判定する。この場合には論理条件成立判定処理ステップ1203が実行される。論理条件が成立したとき、すなわち論理条件（10-2）式では、項カウンタの値が1のとき論理条件が成立していることを示している。

【0088】論理条件が成立した場合、文書識別情報判定処理ステップ1203aが実行され、さらに本文書識別情報が最初の文書識別情報でない場合、結果出力処理ステップ1204を実行する。文書識別情報判定処理ステップ1203aにおいて、最初の文書識別情報の場合、判定処理を行わないが、これは、まだ論理条件の判定処理を何も行っていないためである。結果出力処理ステップ1204では、後述する照合情報を格納した出力バッファ331の内容と論理条件判定結果（以後、論理条件識別情報と呼ぶ）を複合条件判定結果206として出力する。このように、文書識別情報を検出したときに、論理条件判定処理を行うのは、1文書単位に指定論理条件の成否を判定する必要があるためである。

【0089】ここで出力される論理条件判定結果情報の内容としては、位置情報には文書の先頭位置情報および文書の末尾位置情報が設定され、識別子には論理条件識別子Liが格納される照合情報（Li、文書先頭位置情報、文書末尾位置情報）となる。この文書先頭位置情報としては常に文書の先頭で位置情報が0クリアされるため、必ず0（ゼロ）となる。したがって論理条件識別情報（Li、0、文書末尾位置情報）となる。また論理条件成立判定処理ステップ1203の後、初期設定処理ステップ1205を実行し、出力バッファ331を0クリ

アするとともに、要素カウンタと項カウンタの初期設定を行う。本例では要素カウンタに2を設定し、項カウンタに0を設定する。

【0090】照合情報識別処理ステップ1202において照合情報が照合ターム識別情報であると判定された場合には、照合ターム識別処理ステップ1206を実行し、照合ターム識別情報内の照合ターム識別子が論理条件の要素に指定されているか否かを調べる。ここで要素に指定されている場合には、要素判定処理ステップ1206aを実行し、まず見つかった要素について論理条件を調べ、要素に「 $\neg$ 」が掛かっていない場合にはカウンタアップ処理ステップ1206bを実行し、要素カウンタに1を加算する。逆に要素に「 $\neg$ 」が掛かっている場合にはカウンタダウン処理ステップ1206cを実行し、要素カウンタから1を減算する。ここで要素カウンタの値が0となった場合には項の不成立を表し、1以上となった場合には項の成立を表す。項に「 $\neg$ 」が掛かっているときは逆に、要素カウンタの値が0のとき成立を表わし、1以上のとき不成立を表わす。

【0091】次に項成立判定処理ステップ1207を実行する。ここでは要素が指定されている項が不成立から成立に変化したのか、もしくは成立から不成立に変化したのかを調べる。項成立判定方法としては、要素カウンタが、この計算前と計算後で、不成立から成立に変化した場合は成立した項が1つ増えたので項カウンタに1を加算し、成立から不成立に変化した場合は成立した項が1つ減ったため項カウンタから1を減算する。本例では、初期状態における要素カウンタの値は2で、項カウンタの値は0である。ここで要素Taが見つかった場合、要素Taには「 $\neg$ 」が掛かっているため要素カウンタから1を減算するため、要素カウンタの値は1となる。この場合項には「 $\neg$ 」が掛かっているため、項は不成立から不成立へ変化したため、項カウンタは更新しない。さらに要素Tbが見つかった場合、要素Tbには「 $\neg$ 」が掛かっているため要素カウンタから1を減算する。要素カウンタの値は0となり、項は不成立から成立に変化したこととなるため、項カウンタに1を加算する。この結果、項カウンタの値は1となる。以上のようにして、項成立判定処理が行われる。

【0092】照合情報識別処理ステップ1202の処理の後、照合情報退避処理ステップ1210を実行し、照合情報を出力バッファ331に出力する。照合情報退避処理ステップ1210を実行する。この出力バッファ331には1文書分の照合情報が格納され、論理条件成立判定処理ステップ1203で論理条件成立が判明したときに複合条件判定結果206として出力される。以上の各処理ステップをバッファ370に格納されている照合情報に対し繰返し実行していくことにより論理条件判定処理を実現することができる。繰返し処理ステップ1200終了後、最後に処理した文書については論理条件判

定が実行されないことになる。これは、文書識別情報の入力をタイミングとして、前に読み込んだ文書の論理条件判定を行っているからである。したがって、ここで再度、論理条件成立判定処理ステップ1203を実行し、最後に読み込んだ文書についての論理条件判定を行う。

【0093】以上の処理手順を具体例で説明する。検索条件としては、(7-1)に示した式“Q=((文書[4C]理解)[S]システム)[AND](文書[s]検索)”を例に用いて説明する。各条件判定プログラムには検索制御手段101で解析され、各条件に分離された条件式が設定される。具体的には、論理条件判定プログラム330には本条件式(7-1)の論理条件部分“(文書[4C]理解)[S]システム)[AND](文書[s]検索)”が設定される。ここでは、文脈条件“(文書[4C]理解)[S]システム”の識別子をC1とし、文脈条件“文書[s]検索”の識別子をC2とすることにより、(7-1)の論理条件部分を“L1:C1[AND]C2”という形で表す。L1は論理条件識別子である。さらに“L1: $\neg$ ( $\neg$ C1)+( $\neg$ C2)”に変換され、論理条件判定プログラム330に設定される。

【0094】今、(7-2)に示した文書

“...。文書理解を用いた検索システムである。...”

が入力されたとすると、前述したように文脈条件判定プログラム320からは、図22に示す以下の(11-1)~(11-8)が照合情報としてバッファ370へ出力される。

(D1, 0, 0)	(11-1)
(T1, 31, 32)	(11-2)
(T2, 33, 34)	(11-3)
(P1, 31, 34)	(11-4)
(T3, 39, 40)	(11-5)
(T4, 41, 44)	(11-6)
(C1, 30, 48)	(11-7)
(C2, 30, 48)	(11-8)

(11-1)は文書識別情報を表しており、D1は文書識別子を表し、これに続く2項は定数0である。(11-2)、(11-3)、(11-5)、および(11-6)は、それぞれ“文書”、“理解”、“検索”、および“システム”の照合ターム識別情報を表わす。また、T1、T2、T3、およびT4は、それぞれ“文書”、“理解”、“検索”、および“システムの照合ターム識別子”を表わす。また、(11-4)は近傍条件“文書[4C]理解”の近傍条件識別情報を表しており、P1はこの近傍条件識別子を表している。さらに、(11-7)は文脈条件“(文書[4C]理解)[S]システム”の文脈条件識別情報を表しており、C2はこの文脈条件識別子を表し、(11-8)は文脈条件“文書[s]検索”の文脈条件識別情報を表しており、C2はこの文脈

条件識別子を表している。

【0095】以上の条件における論理条件判定処理について図24a、図24bと図25を用いて1ステップずつ説明する。初期状態では図24a、図24bの初期状態に示すようにバッファ370に照合情報(11-1)～(11-8)が格納されており、出力バッファ331及びバッファ370は0クリアされた状態になっている。論理条件判定処理プログラム330ではバッファ370からこれらの照合情報をつづつ読み、論理条件“ $L1: \neg((\neg C1) + (\neg C2))$ ”について判定処理を行う。

【0096】まず図24a、図24bのステップ1に示すように、読み込み処理ステップ1201が実行され照合情報(11-1)、すなわち照合情報(D1, 0, 0)がプログラムのワークエリアに読み込まれる。次に照合情報識別処理ステップ1202が実行され、照合情報(11-1)が文書識別情報であるか照合ターム識別情報であるかが調べられる。照合情報(D1, 0, 0)は後部の2項が両方とも0(ゼロ)のため、文書識別情報と判断される。したがって、論理条件成立判定処理ステップ1203および文書識別情報判定処理ステップ1203aが実行されるが、最初の文書識別情報なので結果出力処理ステップ1204は実行されない。次に、初期設定処理ステップ1205を実行し、出力バッファ331を0クリアし、要素カウンタと項カウンタの初期設定を行う。本例では、要素カウンタの初期値としては1が掛かっている要素( $\neg C1$ )と( $\neg C2$ )の数である2を設定する。また、本項( $(\neg C1) + (\neg C2)$ )は1が掛かっている要素を含む1が掛かっている項である。このため項カウンタの初期値としては、1が掛かっている要素を含まない1が掛かっている項や1が掛かっている要素を含む1が掛かっていない項が存在しないため0を設定する。照合情報識別処理ステップ1202が終了したら、次に照合情報退避処理ステップ1210が実行され、照合情報(D1, 0, 0)が出力バッファ331に格納される。

【0097】その後ステップ2として、再び照合情報読み込み処理ステップ1201が実行され、照合情報(11-2)すなわち(T1, 31, 32)が読み込まれる。次に照合情報識別処理ステップ1202が実行され、照合情報(T1, 31, 32)が照合ターム識別情報かどうか調べられる。照合情報(T1, 31, 32)は下位32ビットが0(ゼロ)でないため、照合ターム識別情報と判断される。そして、次の照合ターム識別処理ステップ1206で、本照合ターム識別情報が論理条件中に指定された要素に該当するかどうか調べられる。本論理条件には、照合情報(T1, 31, 32)の照合識別子T1が指定されていないため、処理ステップ1206a～1209は実行されないことになる。照合情報識別処理ステップ1202が終了した後、照合情報退避処理ステップ

1210が実行され、照合情報(T1, 31, 32)が出力バッファ331に格納される。同様にステップ6まで照合情報の照合情報識別子は本論理条件に指定されていないので照合情報退避処理ステップ1210だけが実行され、照合情報(T2, 33, 34)、(P1, 31, 34)、(T3, 39, 40)、および(T4, 41, 44)が出力バッファ331に格納される。

【0098】さらにステップ7として、再び照合情報読み込み処理ステップ1201が実行され照合情報(11-7)、すなわち(C1, 30, 48)が読み込まれる。次に照合情報識別処理ステップ1202が実行され、照合情報(C1, 30, 48)が照合ターム識別情報かどうか調べられる。照合情報(C1, 30, 48)は下位32ビットが0(ゼロ)でないため、照合ターム識別情報と判断される。そして、次の照合ターム識別処理ステップ1206で、照合ターム識別情報が論理条件中に指定された要素に該当するかどうか調べられる。本論理条件には、照合情報(C1, 30, 48)の照合識別子C1が指定されているため、処理ステップ1206a～1209が実行されることになる。まず、要素判定処理ステップ1206aが実行される。本論理条件では、要素カウンタの初期値は2で項カウンタの初期値は1である。ここで要素C1には1が掛かっているのでカウントダウン処理ステップ1206cが実行される。ここで要素カウンタから1だけ減算する。その結果、要素カウンタの値は2から1に更新される。次に項判定処理ステップ1207が実行される。この時点の要素カウンタの値は1なので、項は不成立から不成立への変化となる。このため、項カウンタは更新されない。照合情報識別処理ステップ1202が終了した後、照合情報退避処理ステップ1210が実行され、照合情報(C1, 30, 48)が出力バッファ331に格納される。

【0099】最後にステップ8として、照合情報読み込み処理ステップ1201が実行され照合情報(11-8)、すなわち(C2, 30, 48)が読み込まれる。次に照合情報識別処理ステップ1202が実行され、照合情報(C2, 30, 48)が照合ターム識別情報かどうか調べられる。照合情報(C2, 30, 48)は下位32ビットが0(ゼロ)でないため、照合ターム識別情報と判断される。そして、次の照合ターム識別処理ステップ1206で、照合ターム識別情報が論理条件中に指定された要素に該当するかどうか調べられる。本論理条件には、照合情報(C2, 30, 48)の照合識別子C2が指定されているため、処理ステップ1206a～1209が実行されることになる。この時点の要素カウンタの値は0で、項カウンタの値は1となっている。まず、要素判定処理ステップ1206aが実行される。ここで、要素C2には1が掛かっているのでカウントダウン処理ステップ1206cが実行される。ここで要素カウンタから1だけ減算する。その結果、要素カウンタの値は1か



ら0となる。次に項判定処理ステップ1207が実行される。このとき本項には、一が掛かっているので不成立から成立への変化が起こったことになる。このため、カウントアップ処理ステップ1208が実行され、項カウンタは1を加算され0から1に更新される。その後、照合情報退避処理ステップ1210が実行され、照合情報(C2, 30, 48)は出力バッファ331に格納される。ここで繰返し処理1200は終了するが、最後の文書なので論理条件成立判定処理ステップ1203が実行される。項カウンタが1なので、本論理条件が成立している。このため、結果出力処理ステップ1204が実行され、照合情報を格納した出力バッファ331の内容と論理条件判定結果情報(L1, 0, 99)が複合条件判定結果206として出力される。L1は論理条件識別子を、0は文書先頭位置情報を、99は文書末尾位置情報を表している。

【0100】以上の論理条件判定処理の結果、最終的には、以下に示す(11-1)～(11-9)の照合情報が、検索条件式(7-1)に示した式“Q=((文書[4C]理解)[S]システム)[AND](文書[s]検索)”の複合条件判定結果206として出力される。

(D1, 0, 0) (11-1)  
(T1, 31, 32) (11-2)  
(T2, 33, 34) (11-3)  
(P1, 31, 34) (11-4)  
(T3, 39, 40) (11-5)  
(T4, 41, 44) (11-6)  
(C1, 30, 48) (11-7)  
(C2, 30, 48) (11-8)  
(L1, 0, 99) (11-9)

(11-1)は文書識別情報を表しており、D1は文書識別子を表わす。(11-2)、(11-3)、(11-5)、および(11-6)は、それぞれ“文書”、“理解”、“検索”、および“システム”の照合ターム識別情報を表わす。また、(11-4)は近傍条件“文書[4C]理解”の近傍条件識別情報を表している。さらに、(11-7)は文脈条件“(文書[4C]理解)[S]システム”の文脈条件識別情報を表し、(11-8)は文脈条件“文書[s]検索”の文脈条件識別情報を表している。最後の(11-9)は論理条件

“(文書[4C]理解)[S]システム)[AND](文書[s]検索)”

の論理条件識別情報を表しており、L1はこの論理条件識別子を表している。以上が論理条件判定プログラム310、文脈条件判定プログラム320、および論理条件判定プログラム330により構成される複合条件判定処理の実現方法である。

【0101】以上説明したように文字列照合回路200および複合条件判定回路300を構成することにより、近傍条件、文脈条件、および論理条件などの複合条件判

定条件を一貫して実現することができるためフルテキストサーチ特有の木目細かな検索が可能となる。さらに、例えば3つのマイクロコンピュータで各々、近傍条件判定プログラム310、文脈条件判定プログラム320、および論理条件判定プログラム330を実行させることにより、プログラム間で同期を取らなくとも動作させることが可能となる。すなわち、これらのプログラムはそれぞれの入力バッファに照合情報が格納されるとこれに応じて条件判定処理を始めるというパイプライン処理を行うことが可能となり、高速な複合条件判定処理を実現することができる。

【0102】次に本発明の第2の実施例について図26を用いて説明する。本実施例では、複合条件のうち1つの条件しか設定されない場合に1つのマイクロプロセッサしか動作しないにもかかわらず、常に3つの複合条件のパイプライン処理を行わなければならないという第1の実施例の欠点を、与えられた検索条件に使用されていない複合条件がある場合にはその複合条件判定プログラムをバイパスすることにより解決し、効率の良い複合条件判定処理を提供するものである。本実施例は文字列照合回路200と複合条件判定回路300aからなり、複合条件判定回路300aは、3つのマイクロコンピュータ、すなわちMPUa301、MPUb302、およびMPUc303と、マルチプレクサ390～392、並びにセクタ380～382から構成されている。マイクロコンピュータMPUa301では近傍条件判定プログラム310が、マイクロコンピュータMPUb302では文脈条件判定プログラム320が、マイクロコンピュータMPUc303では論理条件判定プログラム330が実行される。さらに各MPU間にはファーストイン・ファーストアウト(FIFO)メモリを使用したバッファ350、360、および370が配置され、それぞれのMPU間のデータの受渡しに用いられる。

【0103】本実施例の特徴である複合条件判定回路のパイプライン機能について述べる。本機能は指定された検索条件式で用いられていない複合条件に対応する判定プログラムが搭載されているマイクロプロセッサをマルチプレクサやセクタを用いてバイパスすることにより、使用する複合条件判定プログラムが搭載されているマイクロプロセッサのみを実行させる機能である。例えば検索条件式“文書[4C]理解”が指定された場合には、近傍条件のみが使用されるため、文字列照合回路200から得た照合情報205をバッファ350経由で近傍条件判定プログラム310に入力し、近傍条件判定プログラム310では近傍条件判定を行い、照合情報をバッファ360を経由し、直接複合条件判定結果206として送出するようにする。また、検索条件式

“(文書[S]理解)[AND]システム”が指定された場合には、文脈条件と論理条件が使用されるため、文字列照合回路200から得た照合情報205をバッファ

350経由で直接、文脈条件判定プログラム320に入力し、文脈条件判定プログラム320から出力される照合情報はバッファ370を経由して論理条件判定プログラム330に入力し、この論理条件判定プログラム330から出力される照合情報を複合条件判定結果206として送出するようにする。

【0104】このような複合条件判定回路300aのバイパス機能の具体的な実現方式について説明する。このバイパス動作はマルチプレクサ390~392とセレクト380~382の設定により実現する。この設定情報は第2図の検索制御手段101からの検索情報202として与えられる。マルチプレクサ390では、文字列照合回路200からの照合情報205を、a1が選択された場合は論理条件判定プログラム330へ、b1が指定された場合は文脈条件判定プログラム320へ、c1が指定された場合は近傍条件判定プログラム310へ、d1が選ばれた場合には複合条件判定回路300aの複合条件判定結果206として送出することになる。マルチプレクサ391では、近傍条件判定プログラム310の出力を、a2が指定された場合には文脈条件判定プログラム320へ、b2が指定された場合には論理条件判定プログラム330へ、c2が指定された場合には複合条件判定回路300aの出力として直接送出することになる。マルチプレクサ392では文脈条件判定プログラム320の出力を、a3が指定された場合には論理条件判定プログラム330へ、b3が指定された場合には複合条件判定結果206として送出することになる。

【0105】セレクト380では文脈条件判定プログラム320の入力の選択を行い、マルチプレクサ390にb1が設定されたときはX1を、すなわち文字列照合回路200の照合結果205を選択し、マルチプレクサ391にa2が設定されたときはY1を、すなわち近傍条件判定プログラム310の出力を選択する。セレクト381では論理条件判定プログラム320の入力の選択を行い、マルチプレクサ390にa1が設定されたときはX2を、すなわち文字列照合回路200の照合結果205を選択し、マルチプレクサ391がb2に設定されたときはY2を、すなわち近傍条件判定プログラム310の出力を選択し、マルチプレクサ392がa3に設定されたときはZ2を、すなわち文脈条件判定プログラム320の出力を選択する。セレクト382では複合条件判定結果206の選択を行い、マルチプレクサ390にd1が設定されたときはZ3を、すなわち文字列照合回路200の照合結果205を選択し、マルチプレクサ391がc2に設定されたときはY3を、すなわち近傍条件判定プログラム310の出力を選択し、マルチプレクサ392がb3に設定されたときはX3を、すなわち文脈条件判定プログラム320の出力を選択し、それ以外の場合はW3を選択する、すなわち論理条件判定プログラム330の出力を選択する。以上述べたように、セレクト

380~382の設定は、マルチプレクサ390~392の設定に応じて行うことになる。

【0106】複合条件の組合せにより、以下のようにマルチプレクサの選択を行うことになる。

- ①複合条件なし \_\_\_\_\_  
d1のみを選択
- ②近傍条件のみ \_\_\_\_\_  
c1とc2を選択
- ③文脈条件のみ \_\_\_\_\_  
b1とb3を選択
- ④論理条件のみ \_\_\_\_\_  
a1のみを選択
- ⑤近傍条件と文脈条件 \_\_\_\_\_  
c1, a2, およびb3を選択
- ⑥近傍条件と論理条件 \_\_\_\_\_  
c1とb2を選択
- ⑦文脈条件と論理条件 \_\_\_\_\_  
b1とa3を選択
- ⑧近傍条件と文脈条件と論理条件 —  
c1, a2, およびa3を選択

以上のように、検索制御手段101からの検索情報202として与えられたマルチプレクサ390~392およびセレクト380~382の設定情報に基づき、近傍条件判定プログラム310、文脈条件判定プログラム320、および論理条件判定プログラム330などの複合条件判定プログラムをマルチプレクサ390~392およびセレクト380~382を用いて選択的に接続することにより、複合条件の内1つの条件しか設定されない場合に1つのマイクロプロセッサしか使用しないにもかかわらず、常に3つの複合条件判定プログラムのパイプライン処理を行うという第1の実施例の欠点を解決し、効率の良い複合条件判定処理を実現することが可能となり、高速な検索を行う文書検索装置を実現することができる。

【0107】次に本発明の第3の実施例について図27を用いて説明する。本実施例では、複合条件判定回路300bにおける、近傍条件判定プログラム310、文脈条件判定プログラム320、および論理条件判定プログラム330の3つの複合条件判定プログラムを1つのマイクロコンピュータ上に載せ、これらの複合条件を切り換えて順番に実行させることにより、複合条件判定処理を実現することを目的としている。本実施例の複合条件判定回路300bは、処理速度では第1の実施例に劣るが、1つのマイクロコンピュータで実現できるためコストを低くできるという効果が得られる。

【0108】本実施例は文字列照合処理200と複合条件判定回路300bからなり、さらに複合条件判定回路300bは、マイクロコンピュータMPU301および文字列照合回路200とのデータの受渡しを行うバッファ350から構成される。マイクロコンピュータMP

Ua301には、近傍条件判定プログラム310、文脈条件判定プログラム320、および論理条件判定プログラム330などの3つの複合条件判定プログラムと、これらのプログラムを切り替えるスケジューラ340が搭載されている。ここでバッファ360、370はプログラムのワークエリアとして確保されるが、プログラムによりファーストイン・ファーストアウト（FIFO）メモリとして使用することで、第1の実施例と同様の機能を得ることができる。スケジューラ340における、複合条件プログラムの切替え順序は、まず近傍条件判定プログラム310を実行し、次に文脈条件判定プログラム320、さらに次は論理条件判定プログラム330を実行し、近傍条件判定プログラム310に戻るような順序に複合条件プログラムの切替えを行う。また、スケジューラ340における複合条件プログラムを切り換えるタイミングは、n個の照合ターム情報処理する毎やn件文書処理する毎などにすることも可能である。これはスケジューラ340のプログラムの切替え処理時間との兼ね合いで定めることになる。もし、頻繁に切り替えるとプログラムの切替え時間が複合条件プログラムの実行処理時間に対し、大きな割合を占めるようになるため、数百～数千の照合情報、もしくは数十から数百文書でプログラムを切り替えるのが効果的である。以上のように複合条件判定回路300bを構成することにより、第1の実施例よりも処理速度では劣るものの、1つのマイクロコンピュータで近傍条件判定プログラム310、文脈条件判定プログラム320、および論理条件判定プログラム330などの複合条件判定処理を実現できるため、コストの低い文書検索装置を実現することができる。

【0109】次に本発明の第4の実施例について図28を用いて説明する。本実施例では複合条件判定回路300cにおける、近傍条件判定プログラム310および文脈条件判定プログラム320の2つの複合条件判定プログラムを1つのマイクロコンピュータに載せ、これらの複合条件を切り換えて順番に処理させることにより、近傍条件および文脈の2つの複合条件判定処理を実現させることを目的としている。また、論理条件判定プログラム330は、別のマイクロコンピュータMPUb303に搭載する。本実施例の複合条件判定回路300cは、処理速度では第1の実施例に劣るが、2つのマイクロコンピュータで実現できるためコストを抑えることができるという効果が得られる。本実施例は文字列照合回路200と複合条件判定回路300cからなり、複合条件判定回路300cは、マイクロコンピュータMPUa305およびマイクロコンピュータMPUb303と文字列照合回路200とのデータの受渡しを行うバッファ350、370から構成される。

【0110】マイクロコンピュータMPUaには、近傍条件判定プログラム310および文脈条件判定プログラム320の2つの複合条件判定プログラムと、これらの

プログラムを切り替えるスケジューラ341が搭載されている。ここでバッファ360はプログラムのワークエリアとして確保されるが、プログラムによりファーストイン・ファーストアウト（FIFO）メモリとして使用することにより、第1の実施例と同様の機能を得ることができる。スケジューラ341では、まず近傍条件判定プログラム310を実行し、次に文脈条件判定プログラム320を実行し、その後で近傍条件判定プログラム310に戻るような順序に複合条件プログラムの切替えを行う。また、スケジューラ341における複合条件プログラムを切り換えるタイミングは、n個の照合ターム情報を処理する毎やn文書を処理する毎などにする。これはスケジューラ341のプログラムの切替え処理時間との兼ね合いで定めることになる。もし、頻繁に切り替えるとプログラムの切替え時間が複合条件処理プログラムの実行時間に対し、大きな割合を占めるようになるため、数百～数千の照合情報、もしくは数十から数百文書でプログラムを切り替えるのが効果的である。以上のように複合条件判定回路300cを構成することにより、第1の実施例よりも処理速度では劣るが、2つのマイクロコンピュータで、近傍条件判定プログラム310、文脈条件判定プログラム320、および論理条件判定プログラム330などの複合条件判定処理を実現できるため、コストを抑えた比較的高速な文書検索装置を提供することができる。

【0111】次に本発明の第5の実施例について図29を用いて説明する。本実施例では、第4の実施例において複合条件のうち1つの条件しか設定されない場合でも常に2つのマイクロプロセッサによる複合条件のバイパス処理を行うという欠点を、与えられた検索条件に使用されていない複合条件がある場合にはその複合条件判定プログラムをバイパスすることにより解決し、処理効率の良い複合条件判定回路300dを提供することを目的としている。本実施例は文字列照合回路200と複合条件判定回路300dからなり、複合条件判定回路300dは、マイクロコンピュータMPUa305およびマイクロコンピュータMPUb303と文字列照合回路200とのデータの受渡しを行うバッファ350と370、さらにマルチプレクサ391と392およびセレクタ380と381から構成される。

【0112】本実施例の特徴である複合条件判定回路のバイパス機能について述べる。本機能は指定された検索条件式で使用されていない複合条件判定プログラムが搭載されているマイクロプロセッサをマルチプレクサやセレクタを用いてバイパスすることにより、使用する複合条件判定プログラムが搭載されているマイクロプロセッサのみを実行させる機能である。例えば検索条件式“文書[4C]理解”が指定された場合には、近傍条件のみが使用されるため、文字列照合回路200から得た照合情報205をバッファ350経由で近傍条件判定プログ

ラム310および文脈条件判定プログラム320に入力し、これから出力されて照合情報をバッファ370経由で直接複合条件判定結果206として送出するようにする。また、検索条件式“理解[AND]システム”が指定された場合には、論理条件が使用されているため、文字列照合回路200から得た照合情報205をバッファ350経由で直接、論理条件判定プログラム330に入力し、この論理条件判定プログラム330から出力される照合情報を複合条件判定結果206として送出するようにする。

【0113】このような複合条件判定プログラムのパイパス機能の具体的な実現方法について説明する。このような動作はマルチプレクサ391と392とセクタ380と381の設定により実現する。この設定情報は第2図の検索制御手段101からの検索制御情報202として与えられる。マルチプレクサ391では、文字列照合回路200からの照合情報205を、a2が指定された場合には論理条件判定プログラム330へ送出することになる。また、b2が指定された場合には、文字列照合回路200からの照合情報205を近傍条件判定プログラム310へ送出し、さらにc2が指定された場合には複合条件回路300の出力として直接送出することになる。マルチプレクサ392では文脈条件判定プログラム320の出力を、a3が指定された場合には論理条件判定プログラム330へ、b3が指定された場合には複合条件判定結果206として送出する。

【00114】セクタ380では論理条件判定プログラム330の入力の選択を行い、マルチプレクサ391にa2が設定されたときはX1を、すなわち文字列照合回路200の照合結果205を選択し、マルチプレクサ392にa3が設定されたときはY1を、すなわち文脈条件判定プログラム320の出力を選択することになる。セクタ381では複合条件判定結果206の選択を行い、マルチプレクサ391にc2が設定されたときはX2を、すなわち文字列照合回路200の照合結果205を選択し、マルチプレクサ392がb3に設定されたときはY2を、すなわち文脈条件判定プログラム320の出力を選択し、それ以外の場合はZ2を、すなわち論理条件判定プログラム330の出力を選択する。以上述べたように、セクタ380と381の設定は、マルチプレクサ391と392の設定に応じて行うことになる。

【0115】複合条件の組合せにより、以下のようにマルチプレクサの選択を行うことになる。

- (1) 複合条件なし \_\_\_\_\_  
c2のみを選択
- (2) 論理条件のみ \_\_\_\_\_  
a2のみを選択
- (3) 近傍条件と文脈条件 \_\_\_\_\_  
b2とb3を選択

(4) 近傍条件と文脈条件と論理条件 —

b2とa3を選択

以上のように、検索制御手段101からの検索制御情報202として与えられたマルチプレクサ391と392およびセクタ381と382の設定情報に基づき、近傍条件判定プログラム310、文脈条件判定プログラム320、および論理条件判定プログラム330などの複合条件判定プログラムをマルチプレクサ391と392およびセクタ380と381を用いて選択的に接続することにより、第4の実施例が複合条件の内1つの条件しか設定されない場合には1つのマイクロプロセッサしか使用されないにもかかわらず、常に2つマイクロプロセッサにより複合条件のパイプライン処理を行うという欠点を解決し、効率の良い複合条件判定処理を実現することが可能となり、比較的低コストで且つ比較的高速な文書検索装置を実現することができる。

【0116】次に本発明の第6の実施例について図30を用いて説明する。本実施例では複合条件判定回路300eにおける、文脈条件判定プログラム320および論理条件判定プログラム330の2つの複合条件判定プログラムを1つのマイクロコンピュータに搭載し、これらの複合条件を切り換えて順番に処理させることにより、文脈条件および論理条件の2つの複合条件判定処理を1つのマイクロコンピュータで実現することを目的としている。また、近傍条件判定プログラム310は、別のマイクロコンピュータに受け持たせる。本実施例の複合条件判定回路300eは、処理速度では第1の実施例に劣るが、2つのマイクロコンピュータで実現できるため比較的高性能を維持してコストを抑えることができるという効果が得られる。

【0117】本実施例は文字列照合回路200と複合条件判定回路300eからなり、複合条件判定回路300eは、マイクロコンピュータMPUa301およびマイクロコンピュータMPUb306と文字列照合回路200とのデータの受渡しを行うバッファ350、360から構成される。マイクロコンピュータMPUa306には、文脈条件判定プログラム320および論理条件判定プログラム330の2つの複合条件判定プログラムと、これらのプログラムを切り替えるスケジューラ342が搭載されている。ここでバッファ370はプログラムのワークエリアに確保されるが、プログラムでファーストイン・ファーストアウト(FIFO)メモリとして使用することにより、第1の実施例と同様の機能を得ることができる。スケジューラ342では、まず文脈条件判定プログラム320を実行し、次に論理条件判定プログラム330を実行し、その後で文脈条件判定プログラム320に戻るといった順序に複合条件プログラムの切替えを行う。また、スケジューラ342における複合条件プログラムを切り換えるタイミングは、n個の照合タム情報を処理する毎やn件の文書を処理する毎などに

することも可能である。これはスケジューラ342のプログラムの切替え処理時間との兼ね合いで定めることになる。もし、頻繁に切り替えるとプログラムの切替え時間が複合条件処理プログラムの実行時間に対し、大きな割合を占めるようになるため、数百～数千の照合情報、もしくは数十から数百文書でプログラムを切り替えるのが効果的である。以上のように複合条件判定回路300eを構成することにより、第1の実施例よりも処理速度では劣るが、2つのマイクロコンピュータで、近傍条件判定プログラム310、文脈条件判定プログラム320、および論理条件判定プログラム330などの複合条件判定処理を行うため、低コストで比較的高速な文書検索装置を実現することができる。

【0118】次に本発明の第7の実施例について図31を用いて説明する。本実施例では、第6の実施例において複合条件のうち1つの条件しか設定されない場合でも常に2つのマイクロプロセッサによる複合条件のバイパス処理を行うという問題点を、与えられた検索条件に使用されていない複合条件がある場合にはその複合条件判定プログラムをバイパスすることにより解決し、処理効率の良い複合条件判定回路300fを実現することを目的としている。本実施例は文字列照合回路200と複合条件判定回路300fからなり、複合条件判定回路300fは、マイクロコンピュータMPUa301およびマイクロコンピュータMPUb306と文字列照合回路200とのデータの受渡しを行うバッファ350と360、さらにマルチプレクサ391と392およびセクタ380と381から構成される。

【0119】本実施例の特徴である複合条件判定回路のバイパス機能について述べる。本機能は指定された検索条件式に使用されていない複合条件判定プログラムが搭載されているマイクロプロセッサをマルチプレクサやセクタを用いてバイパスすることにより、使用する複合条件判定プログラムが搭載されているマイクロプロセッサのみを実行させる機能である。例えば検索条件式“文書[4C]理解”が指定された場合には、近傍条件のみが使用されるため、文字列照合回路200から得た照合情報205をバッファ350経由で近傍条件判定プログラム310に入力し、この出力の照合情報をバッファ370経由で直接複合条件判定結果206として送出する。また、検索条件式“理解[AND]システム”が指定された場合には、論理条件が使用されているため、文字列照合回路200から得た照合情報205をバッファ350経由で直接、文脈条件判定プログラム320に入力し、さらにバッファ370を経由し論理条件判定プログラム330に入力し、この論理条件判定プログラム330の出力の照合情報を複合条件判定結果206として送出する。

【0120】このような複合条件判定回路のバイパス機能の具体的な実現方法について説明する。このような動

作はマルチプレクサ391と392とセクタ380と381の設定により実現される。この設定情報は第2図の検索制御手段101からの検索情報202として与えられる。マルチプレクサ391では、文字列照合回路200からの照合情報205の出力を、a2が指定された場合には文脈条件判定プログラム320へ、b2が指定された場合には近傍条件判定プログラム310へ、c2が指定された場合には複合条件回路300の出力として直接送出することになる。マルチプレクサ392では近傍条件判定プログラム310の出力を、a3が指定された場合には文脈条件判定プログラム320へ、b3が指定された場合には複合条件判定結果206として送出することになる。

【0121】セクタ380では文脈条件判定プログラム320の入力の選択を行い、マルチプレクサ391にa2が設定されたときはX1を、すなわち文字列照合回路200の照合結果205を選択し、マルチプレクサ392にa3が設定されたときはY1を、すなわち近傍条件判定プログラム310の出力を選択する。セクタ381では複合条件判定結果206の選択を行い、マルチプレクサ391にc2が設定されたときはX2を、すなわち文字列照合回路200の照合結果205を選択し、マルチプレクサ392がb3に設定されたときはY2を、すなわち近傍条件判定プログラム310の出力を選択し、それ以外の場合はZ2を、すなわち論理条件判定プログラム330の出力を選択する。

【0122】以上述べたように、セクタ380と381の設定は、マルチプレクサ391と392の設定に応じて行うことになる。すなわち複合条件の組合せにより、以下のようにマルチプレクサの選択を行うことになる。

- (1) 複合条件なし \_\_\_\_\_  
c2のみを選択
- (2) 近傍条件のみ \_\_\_\_\_  
b2とb3を選択
- (3) 文脈条件と論理条件 \_\_\_\_\_  
a2のみを選択
- (4) 近傍条件と文脈条件と論理条件 —  
b2とa3を選択

以上のように、検索制御手段101からの検索制御情報202として与えられたマルチプレクサ391と392およびセクタ381と382の設定情報に基づき、近傍条件判定プログラム310、文脈条件判定プログラム320、および論理条件判定プログラム330などの複合条件判定プログラムをマルチプレクサ391と392およびセクタ380と381を用いて選択的に接続することにより、第6の実施例が複合条件の内1つの条件しか設定されない場合には1つのマイクロプロセッサしか使用されないにもかかわらず、常に2つマイクロプロセッサにより複合条件のバイパス処理を行うという

欠点を解決し、効率の良い複合条件判定処理を実現することが可能となり、コストを抑えた比較的高速な文書検索装置を提供することができる。

【0123】次に本発明の第8の実施例について図32を用いて説明する。第1の実施例では、文字列照合回路200から出力される照合情報の中に近傍条件の処理対象にならない文脈識別文字列、すなわち文脈マーカの照合情報も入っているため、近傍条件判定プログラム310では文脈マーカの照合情報についても近傍条件判定処理を実施することになり、近傍条件判定の処理速度が落ちるという問題がある。本実施例では上記問題点を解決する複合条件判定方法として、文脈マーカの照合情報のみを格納するバッファ380を設け、近傍条件判定処理をバイパスし文脈マーカの照合情報が必要な文脈条件判定プログラム320に入力することにより高速な複合条件判定処理が可能な複合条件判定回路300gを実現することが目的である。

【0124】本実施例は文字列照合回路200と複合条件判定回路300gからなり、複合条件判定回路300gは、マイクロコンピュータMPUa301、マイクロコンピュータMPUb302aおよびマイクロコンピュータMPUc303とこれらの間のデータの受渡しを行うバッファ350、360、370および380、さらにマルチプレクサ710、文脈マーカ検出器720から構成される。マルチプレクサ710は通常はポートaを選択し、文字列照合回路200から送られる照合情報205を近傍条件判定プログラム310の入力となるバッファ350へ送出する。また、後述する文脈マーカ検出器720から文脈マーカ検出信号721が送られるとマルチプレクサ710ではポートbを選択し、後述するソートマージプログラム730の入力となるバッファ380へ送出する。すなわち、文脈マーカの照合情報はバッファ380へ送出されることになる。さらに、後述する文脈マーカ検出器720から文書識別情報検出信号722が送られるとマルチプレクサ710ではポートaとポートbの両方を選択し、バッファ350とバッファ380へ同時に送出する。すなわち、文書識別情報はバッファ350とバッファ380の両方へ送出されることになる。

【0125】文脈マーカ検出器720では、文字列照合回路200の出力を参照して、文脈マーカの照合情報か否かを判定する。すなわち、照合情報の照合情報識別子があらかじめ定められた文脈マーカの照合情報識別子と同じもので、且つ先頭位置情報と末尾位置情報とも0（ゼロ）でないということであれば、文脈マーカであると判定する。また、先頭位置情報と末尾位置情報とが共に0（ゼロ）であれば、文書識別情報であると判定する。文脈マーカ検出器720は、文書識別情報検出用のコンパレータ、文脈マーカの照合ターム情報識別用の2つのコンパレータ、文脈マーカの識別子格納

用のレジスタ、および0（ゼロ）を格納するレジスタから構成される。まず、文書識別情報検出用のコンパレータは、照合情報の位置情報が0かどうかを調べ、0の場合には文書識別情報検出信号722を出力する。すなわち、0が格納されているレジスタと照合情報の位置情報を比較し、等しい場合に文書識別情報検出信号722を出力することになる。次に、文脈マーカの照合ターム情報識別用のコンパレータは、文脈マーカの照合ターム情報識別子かどうかを調べるコンパレータと、今調べている照合情報が照合ターム情報かどうかを調べるコンパレータを用意する。ここで、これらの両方のコンパレータから成立信号が出力された場合にのみ文脈マーカ検出信号721を出力する。文脈マーカの照合ターム情報識別子かどうかを調べるコンパレータでは、照合情報の照合情報識別子が文脈マーカの識別子と同じ場合、すなわち照合情報の照合情報識別子と文脈マーカの識別子格納用のレジスタとを比較し、等しい場合に成立信号を出力する。照合ターム情報かどうかを調べるコンパレータでは、照合情報の位置情報が0でない場合、すなわち照合情報の位置情報と0が格納されているレジスタとを比較し、等しくない場合に成立信号を出力する。以上のように、文脈マーカ検出器720を構成する。

【0126】文脈マーカの照合情報が文脈マーカ検出器720に入力された場合、マルチプレクサ710に文脈マーカ検出信号721を出力する。これに応じて、マルチプレクサ710では、文字列照合回路200から送られた照合情報205の送出先をバッファ350からバッファ380へ切り替えることになる。文書識別情報であればマルチプレクサ710に文書識別情報検出信号722を出力する。これに応じて、マルチプレクサ710では、文字列照合回路200から送られた照合情報205の送出先をバッファ350とバッファ380の両方に設定する。

【0127】マイクロコンピュータ302aでは、ソートマージプログラム730と文脈条件判定プログラム320とこれらを制御するスケジューラ342が実行される。ソートマージプログラム730ではバッファ360に格納された近傍条件判定プログラム310の出力としての照合情報と、バッファ380に格納された文脈マーカ照合情報とを、末尾位置情報の昇順にマージする。すなわち、バッファ360とバッファ380から各々照合情報を1つずつ読み込み、これらの末尾位置情報を比較する。そして、バッファ360から読み込んだ照合情報の末尾位置情報がバッファ380から読み込んだ照合情報の末尾位置情報より小さい場合には、バッファ360から読み込んだ照合情報をバッファ390へ先に出力し、次はバッファ360から照合情報を読み込み、同様に先程のバッファ380から読み込んだ照合情報と比較し、小さい方をバッファ390へ出力する。逆にバッ

55

ァ380の照合情報の末尾位置情報がバッファ360の照合情報の末尾位置情報より小さい場合はバッファ380の照合情報をバッファ390へ出力し、次はバッファ380から照合情報を読み込み、同様に先程のバッファ360から読み込んだ照合情報と比較し、小さい方をバッファ390へ出力する。また、文書識別情報はバッファ360から読み込んだもののみをバッファ390に出力し、バッファ380から読み込んだ文書識別情報はバッファ390に出力しない。さらに、ここで行うソートマージは文書毎に処理を行う。

【0128】これらの処理を行うことによりバッファ390には、バッファ360とバッファ380の照合情報が末尾位置情報でソートマージされることになり、第1の実施例における近傍条件判定プログラム310の出力した照合情報と同様のものが格納されることになる。スケジューラ343における、プログラムの切替え順序は、まずソートマージプログラム730を実行し、次に文脈条件判定プログラム320を実行し、ソートマージプログラム730に戻るような順序にプログラムの切替えを行う。また、スケジューラ343におけるプログラムを切り換えるタイミングは、n個の照合ターム情報処理する毎やn件文書を処理する毎などにすることも可能である。これはスケジューラ343のプログラムの切替え処理時間との兼ね合いで定めることになる。もし、頻繁に切り替えるとプログラムの切替え時間がプログラムの実行処理時間に対し、大きな割合を占めるようになるため、数百～数千の照合情報、もしくは数十から数百文書でプログラムを切り替えるのが効果的である。以上のように複合条件判定回路300gを実現することにより近傍条件判定処理に必要な文脈マーカの照合情報を近傍条件判定プログラム310をバイパスすることが可能となり、第1の実施例より高速な文書検索装置が実現できる。

【0129】次に本発明の第9の実施例について図33を用いて説明する。本実施例は第8の実施例と同様に、文字列照合回路200から出力される照合情報の中に近傍条件の処理対象にならない文脈マーカの照合情報も入っているため、近傍条件判定プログラム310では文脈マーカの照合情報についても近傍条件判定処理を実施することになり、近傍条件判定の処理速度が落ちるという第1の実施例の問題点を、文脈マーカの照合情報のみを格納するバッファを設け、近傍条件判定処理をバイパスし文脈マーカの照合情報が必要な文脈条件判定プログラム320のみに入力することにより高速な複合条件判定処理が行える複合条件判定回路300hを実現することが目的である。

【0130】本実施例が第8の実施例と異なるのは、文脈マーカ検出専用文脈マーカ用文字列照合回路200aを設けた点である。第8の実施例では文脈マーカを含む検索タームを全て文字列照合回路200に設定

56

するため、文字列照合回路200に設定する検索ターム数が多くなるという問題がある。さらには文字列照合回路200の許容する検索ターム数を超えてしまう場合も出てくる。また、文脈マーカの検索タームは1度設定すれば済むものであるのに対して、第8の実施例のように検索条件が与えられる度に文脈マーカの検索タームも再設定すると、検索情報202の作成時間および設定時間が長くなるという問題点も生じる。

【0131】本実施例は文字列照合回路200と文脈マーカ用文字列照合回路200a、および複合条件判定回路300hから構成される。複合条件判定回路300hは、マイクロコンピュータMPUa301、マイクロコンピュータMPUb302aおよびマイクロコンピュータMPUc303とこれらの間のデータの受渡しを行うバッファ350、360、370および380から構成される。文脈マーカ用文字列照合回路200aの構成は文字列照合回路200と同様であり、文脈マーカ用文字列照合回路200aには文脈マーカを検索タームとして設定し、文字列照合回路200には文脈マーカ以外の検索タームを設定する。また、検索制御手段101では、文字列照合回路200には検索条件が与えられる度に検索情報202を設定するが、文脈マーカ用文字列照合回路200aは本検索装置の立上時に1度だけ設定する。

【0132】本実施例の処理手順を具体例で説明する。まず、文字列照合回路200と文脈マーカ用文字列照合回路200aの動作を具体例で説明する。(7-1)に示した式

“Q=((文書[4C]理解)[S]システム)[AND](文書[S]検索)”

を例に用いて説明する。各複合条件判定プログラムには検索制御手段101で解析され、各条件に分離された条件式が設定される。本例では検索制御手段101より文字列照合回路200には、“T1:文書”、“T2:理解”、“T3:検索”、および“T4:システム”の4つが検索タームとして設定され、文脈マーカ用文字列照合回路200aには、“S1:。”が検索タームとして設定される。

【0133】今、(7-2)に示した文書  
“・・・文書理解を用いた検索システムである。・・・”

が入力されたとすると、文字列照合回路200からは以下の照合情報(13-1)～(13-5)が照合情報205としてバッファ350へ出力される。

(D1, 0, 0) (13-1)

(T1, 31, 32) (13-2)

(T2, 33, 34) (13-3)

(T3, 39, 40) (13-4)

(T4, 41, 44) (13-5)

また、文脈マーカ用文字列照合回路200aからは以

下の文脈マーカーの照合情報(12-1)~(12-3)が照合情報205aとしてバッファ380へ出力される。

(D1, 0, 0) (12-1)

(S1, 30, 30) (12-2)

(S1, 48, 48) (12-3)

上記のバッファ360, 380の照合情報は第8の実施例と同じように、近傍条件判定プログラム310、ソートマージプログラム730、文脈条件判定プログラム320、および論理条件判定プログラム330により処理される。以上のように文脈マーカー検出専用の文字列照合回路200aを設けることにより、近傍条件判定処理に必要な文脈マーカーの照合情報を近傍条件判定プログラム310をバイパスすることが可能となり、さらに第8の実施例よりも文脈マーカーに関する検索情報202の作成時間および設定時間が少なくて済むため、第1の実施例より高速な文書検索装置が実現できる。

【0134】次に本発明の第10の実施例について図34を用いて説明する。本実施例は第1の実施例の複合条件判定回路300では、マイクロコンピュータMPUa301、マイクロコンピュータMPUb302、およびマイクロコンピュータMPUc303のデータの受渡しに使用しているバッファ360およびバッファ370にファーストイン・ファーストアウト(FIFO)メモリという特殊なメモリを使用しているため、メモリ容量当りのコストが高く掛かるという問題点がある。本実施例では、この代りに一般のメモリを使用することにより低コストの複合条件判定回路300iを実現することを目的としている。

【0135】本実施例はマイクロコンピュータMPUa301、マイクロコンピュータMPUb302、マイクロコンピュータMPUc303、バッファ350、バス630、および共有メモリ620から構成される。また、マイクロコンピュータMPUa301では近傍条件判定プログラム310が、マイクロコンピュータMPUb302では文脈条件判定プログラム320が、マイクロコンピュータMPUc303では論理条件判定プログラム330が実行される。

【0136】共有メモリ620は近傍条件判定プログラム310、文脈条件判定プログラム320、および論理条件判定プログラム330のデータの受渡しに使用する。すなわち、近傍条件判定プログラム310と文脈条件判定プログラム320とのデータの受渡しには共有メモリ620内のバッファ360aが、文脈条件判定プログラム320と論理条件判定プログラム330とのデータの受渡しにはバッファ370aがそれぞれ使用される。バッファ360aとバッファ370aは、近傍条件判定プログラム310、文脈条件判定プログラム320、および論理条件判定プログラム330においてプログラムでファーストイン・ファーストアウト(FIF

O)メモリとして使用することにより、それぞれバッファ360とバッファ370と同様の機能を得ることが可能である。

【0137】本実施例の複合条件判定処理の動作について説明する。文字列照合回路200の照合結果は照合情報205としてバッファ350に送出される。バッファ350に格納された照合情報は近傍条件判定プログラム310で処理され判定結果は共有メモリ620内のバッファ360aに格納される。次にバッファ360aに照合情報が格納されると文脈条件判定プログラム320が実行され、文脈条件判定プログラム320の判定結果は共有メモリ620内のバッファ370aに格納される。さらにバッファ370aに照合情報が格納されると論理条件判定プログラム330が実行され、論理条件判定プログラム330の判定結果は複合条件判定結果206として送出される。以上のように、複合条件判定回路300iを実現することにより、FIFOメモリの替わりに低コストの通常のメモリをバッファ360a, 370aに使用することが可能となり、低コストの安い文書検索装置を実現することができる。

【0138】最後に本発明の第11の実施例について図35を用いて説明する。第1の実施例の複合条件判定回路300では、マイクロコンピュータMPUa301、マイクロコンピュータMPUb302、およびマイクロコンピュータMPUc303間のデータの受渡しに使用するバッファ360およびバッファ370にファーストイン・ファーストアウト(FIFO)メモリという特殊なメモリを使用しているため、メモリ容量当りのコストが高くつくという問題点がある。本実施例ではこのFIFOメモリの代りに、一般のメモリを使用することにより低コストの複合条件判定回路300jを実現することを目的としている。また、第10の実施例では、共有メモリ620を3つのマイクロプロセッサで時分割でアクセスするため、各プロセッサが同じ回数のメモリアクセスを行うとすると、メモリのアクセス回数は3倍になり処理速度がメモリのアクセスネックになるという問題がある。これに対して本実施例では、各プロセッサ間に一般のメモリを使用したバッファを2面設けることにより解決することを目的としている。

【0139】本実施例はマイクロコンピュータMPUa301、マイクロコンピュータMPUb302、マイクロコンピュータMPUc303、バッファ350、バッファ360bおよびバッファ370bから構成される。また、マイクロコンピュータMPUa301では近傍条件判定プログラム310が、マイクロコンピュータMPUb302では文脈条件判定プログラム320が、マイクロコンピュータMPUc303では論理条件判定プログラム330が実行される。

【0140】バッファ360bは近傍条件判定プログラム310と文脈条件判定プログラム320のデータの受



渡しに使用され、バッファ370bは文脈条件判定プログラム320と論理条件判定プログラム330のデータの受渡しに使用されている。バッファ360bでは、マルチプレクサ630、セクタ631、メモリ622、623、バス640、641、および通信メモリ624から構成されている。バッファ370bもバッファ360bと同様に構成される。2面バッファを構成するメモリ622とメモリ623は近傍条件判定プログラム310から文脈条件判定プログラム320への照合情報の受渡しに使用される。近傍条件判定プログラム310がメモリ622に照合情報を出力している間、文脈条件判定プログラム320はメモリ623から以前に近傍条件判定プログラム310が出力した照合情報を読み込む。また、近傍条件判定プログラム310がメモリ623に照合情報を出力している間、文脈条件判定プログラム320はメモリ622から以前に近傍条件判定プログラム310が出力した照合情報を読み込む。通信メモリ624は、メモリ622およびメモリ623のバッファの切り替えのための制御情報の受渡しに使用される。

【0141】マルチプレクサ630は切り替え信号630aに0が設定されるとポートaを選択し、近傍条件判定プログラム310の出力する照合情報がメモリ622に格納される。切り替え信号630aに1が設定されるとポートbを選択し、近傍条件判定プログラム310の出力する照合情報がメモリ623に格納される。セクタ631は切り替え信号631aに0が設定されるとポートxが選択され、メモリ622から照合情報が文脈条件判定プログラム320により読込まれる。切り替え信号631aに1が設定されるとポートyが選択され、メモリ622から照合情報が文脈条件判定プログラム320により読込まれる。

【0142】以下、バッファ360bの2面バッファ方式の動作の制御方法について説明する。まず、近傍条件判定プログラム310が切り替え信号630aとして0をマルチプレクサ630に送る、すなわち近傍条件判定プログラム310から出力される照合情報をメモリ622に出力する。近傍条件判定プログラム310がメモリ622に所定の量の照合情報を書き込み終えたとき通信メモリ624を経由し、文脈条件判定プログラム320にメモリ622が使用できることを知らせる。文脈条件判定プログラム320はこれを受け、セクタ631に切り替え信号631aとして0を設定する。すなわち、文脈条件判定プログラム320の入力としてメモリ622を選択することになり、メモリ622の照合情報を読み込み文脈条件判定処理を行う。

【0143】次に、近傍条件判定プログラム310が切り替え信号630aとして1をマルチプレクサ630に送る。すなわち近傍条件判定プログラム310から出力される照合情報をメモリ623に出力することになる。近傍条件判定プログラム310がメモリ623に所定の

量の照合情報を書き込み終えたとき通信メモリ624を経由し、文脈条件判定プログラム320にメモリ623が使用できることを知らせる。文脈条件判定プログラム320はこれを受け、セクタ631に切り替え信号631aとして1を設定する。すなわち、文脈条件判定プログラム320の入力としてメモリ623を選択することになり、メモリ623の照合情報を読み込み文脈条件判定処理を行う。

【0144】その後再び、近傍条件判定プログラム310が切り替え信号630aとして0をマルチプレクサ630に送る。このとき、近傍条件判定プログラム310がメモリ622に照合情報を書き込む場合、文脈条件判定プログラム320からのメモリ622の読込み終了が報告されるまで、近傍条件判定プログラム310はメモリ622への照合情報の書き込みを待つことになる。このため文脈条件判定プログラム320においてメモリ622の読込みが終了したとき通信メモリ624を経由し、メモリ622の読込みが終了したことを知らせる。このようにメモリ622およびメモリ623のバッファの切り替えを制御することにより、近傍条件判定プログラム310と文脈条件判定プログラム320とが同じメモリにアクセスすることのない2面バッファ方式が実現される。文脈条件判定プログラム320と論理条件判定プログラム330とのデータの受渡しに使用するバッファ370bもバッファ360bと同様のものを使用することができる。以上のように、実現することにより、FIFOメモリの替わりに低コストの通常のメモリをバッファ360b、370bを用いて複合条件判定回路300jを構成することにより、低コストで、且つ高速な文書検索装置を実現することができる。

#### 【0145】

【発明の効果】以上のように本発明が提供する文書検索方法および装置によれば、複合条件である近傍条件、文脈条件、および論理条件を容易に判定することができ、しかも高速に判定処理することが可能となり、フルテキストサーチ特有の木目細かな検索を高速に実現する文書検索装置を提供することができる。

#### 【図面の簡単な説明】

【図1】本発明を用いた複合条件判定回路の説明図である。

【図2】文書検索装置の説明図である。

【図3】複合条件の判定例の説明図である。

【図4】照合情報の説明図である。

【図5】複合条件の判定例の説明図である。

【図6】複合条件の判定例の説明図である。

【図7】複合条件の判定例の説明図である。

【図8】文書識別子検出回路の説明図である。

【図9】検索ターム長テーブルの説明図である。

【図10】位置情報付加回路の説明図である。

【図11】文書識別情報の説明図である。

【図12】照合ターム識別情報の説明図である。  
 【図13】照合情報の説明図である。  
 【図14】照合情報の説明図である。  
 【図15】近傍条件判定処理の説明図である。  
 【図16a】近傍条件判定処理の説明図の一部である。  
 【図16b】近傍条件判定処理の説明図の他部である。  
 【図17】近傍条件判定処理の説明図である。  
 【図18】文脈条件判定処理の説明図である。  
 【図19】文脈条件判定処理の説明図である。  
 【図20a】文脈条件判定処理の説明図の一部である。  
 【図20b】文脈条件判定処理の説明図の他部である。  
 【図21a】文脈条件判定処理の説明図の一部である。  
 【図21b】文脈条件判定処理の説明図の他部である。  
 【図22】文脈条件判定処理の説明図である。  
 【図23】論理条件判定処理の説明図である。  
 【図24a】論理条件判定処理の説明図の一部である。  
 【図24b】論理条件判定処理の説明図の他部である。  
 【図25】論理条件判定処理の説明図である。  
 【図26】本発明を用いた複合条件判定回路の説明図である。  
 【図27】本発明を用いた複合条件判定回路の説明図である。  
 【図28】本発明を用いた複合条件判定回路の説明図である。  
 【図29】本発明を用いた複合条件判定回路の説明図である。  
 【図30】本発明を用いた複合条件判定回路の説明図である。

ある。

【図31】本発明を用いた複合条件判定回路の説明図である。

【図32】本発明を用いた複合条件判定回路の説明図である。

【図33】本発明を用いた複合条件判定回路の説明図である。

【図34】プロセッサ間のバッファの説明図である。

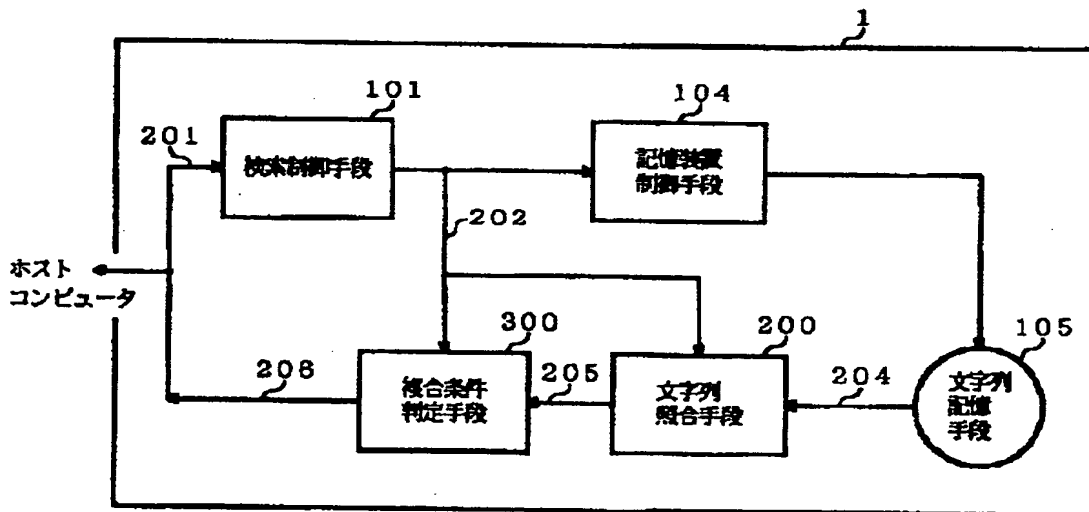
【図35】プロセッサ間のバッファの説明図である。

# 【符号の説明】

101 検索制御手段  
 104 記憶装置制御手段  
 105 文字列記憶手段  
 200 文字列照合手段  
 300 複合条件判定手段  
 210 文書識別子検出回路  
 220 タームコンパレータ  
 230 文字数カウンタ  
 240 位置情報付加回路  
 250 検索ターム長テーブル  
 310 近傍条件判定プログラム  
 320 文脈条件判定プログラム  
 330 論理条件判定プログラム  
 350 バッファ  
 360 バッファ  
 370 バッファ  
 380 バッファ

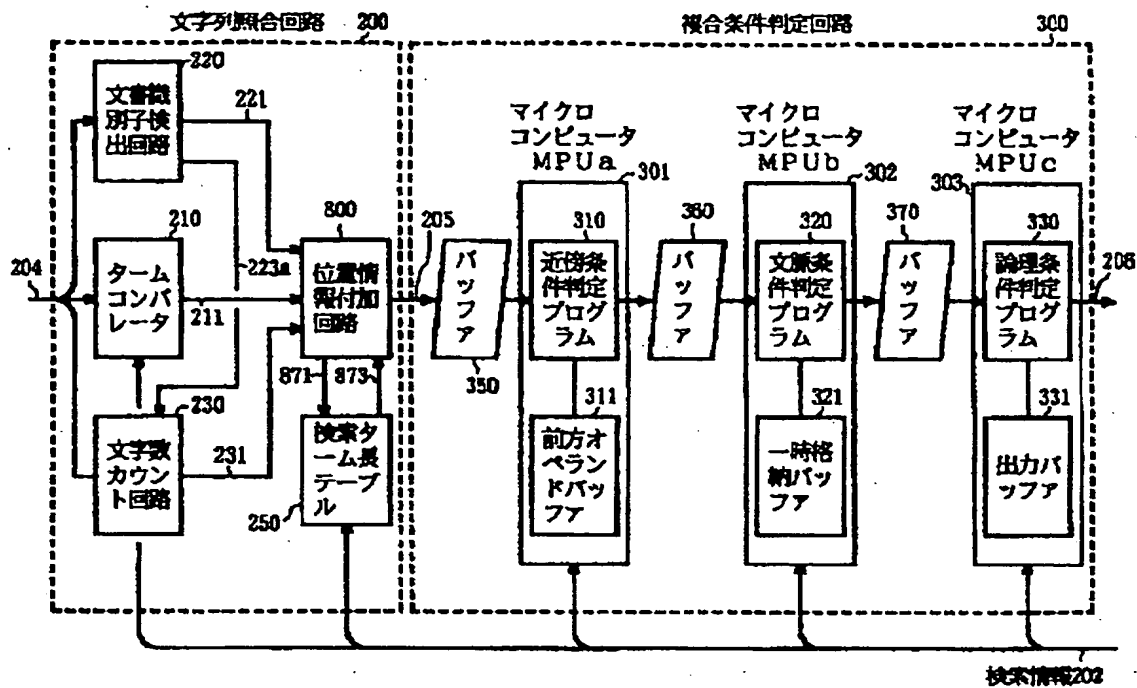
【図2】

【図2】



【図1】

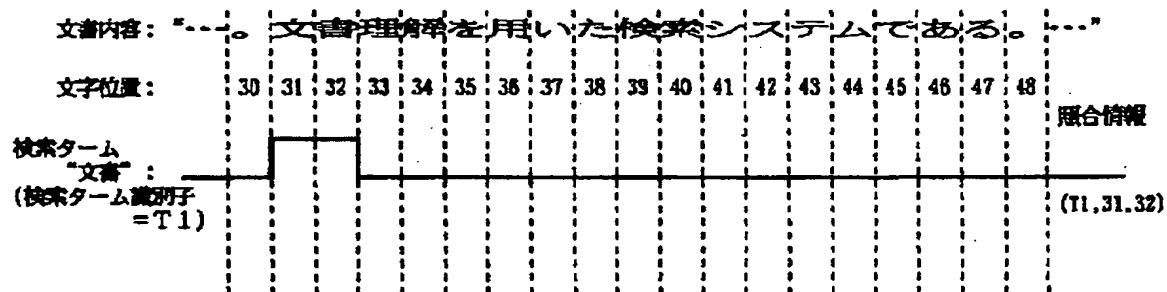
【図1】



【図4】

【図4】

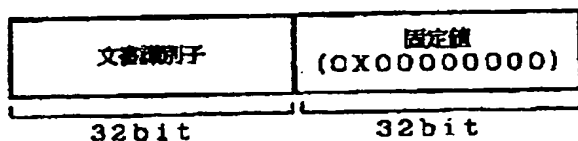
&lt;検索タームの照合情報&gt;



【図11】

【図11】

&lt;文書識別情報&gt;



【図3】

【図3】

検索ターム1: "文書"  
 検索ターム2: "検索"

## ①文書例1

文書内容: 文書の内容を読んで検索する。……………  
 7文字

## ②文書例2

文書内容: 文書の内容を理解して検索を行う。……………  
 8文字

## ③文書例3

文書内容: 検索の結果に基づいて該当文書の内容を表示する。……  
 10文字

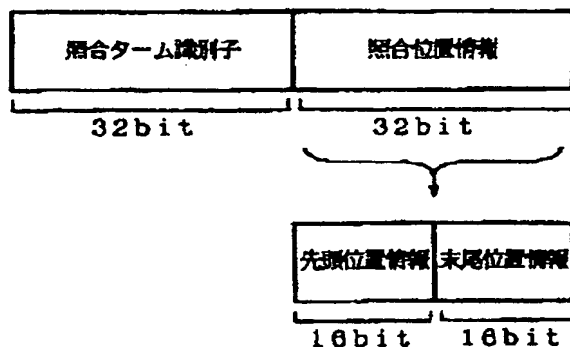
## ④文書例4

文書内容: 文書の論理的な構造を意図した検索を行う。……………  
 12文字

【図12】

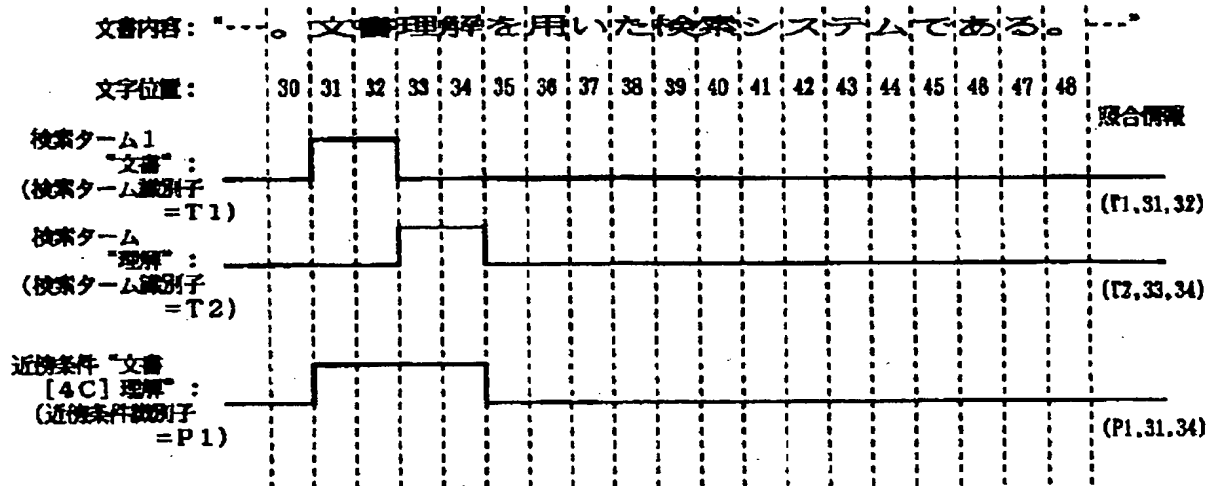
【図12】

〈照合ターム識別情報〉



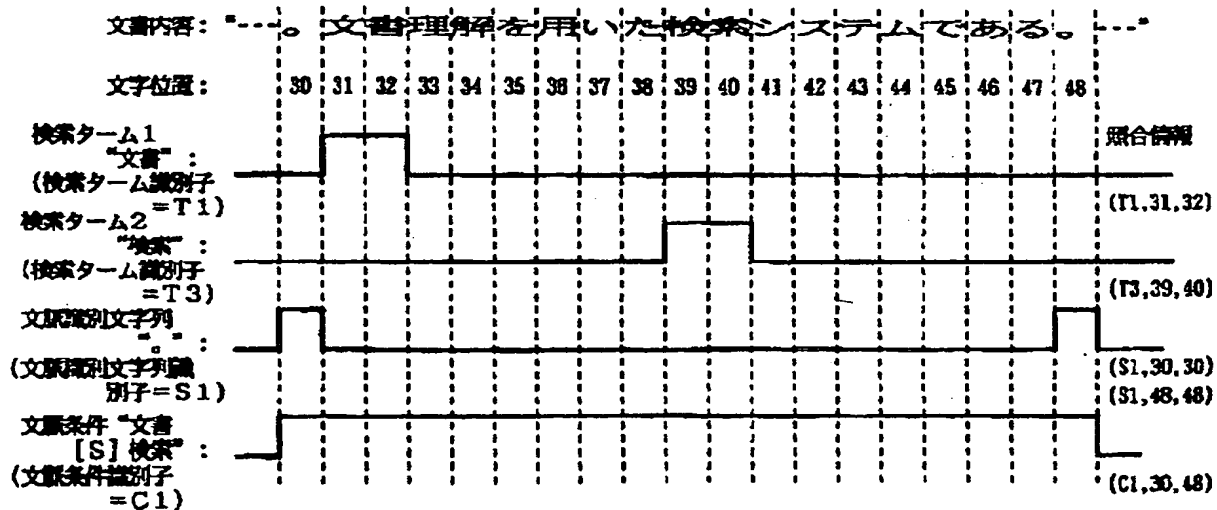
【図5】

【図5】  
 <近傍条件の判定原理>

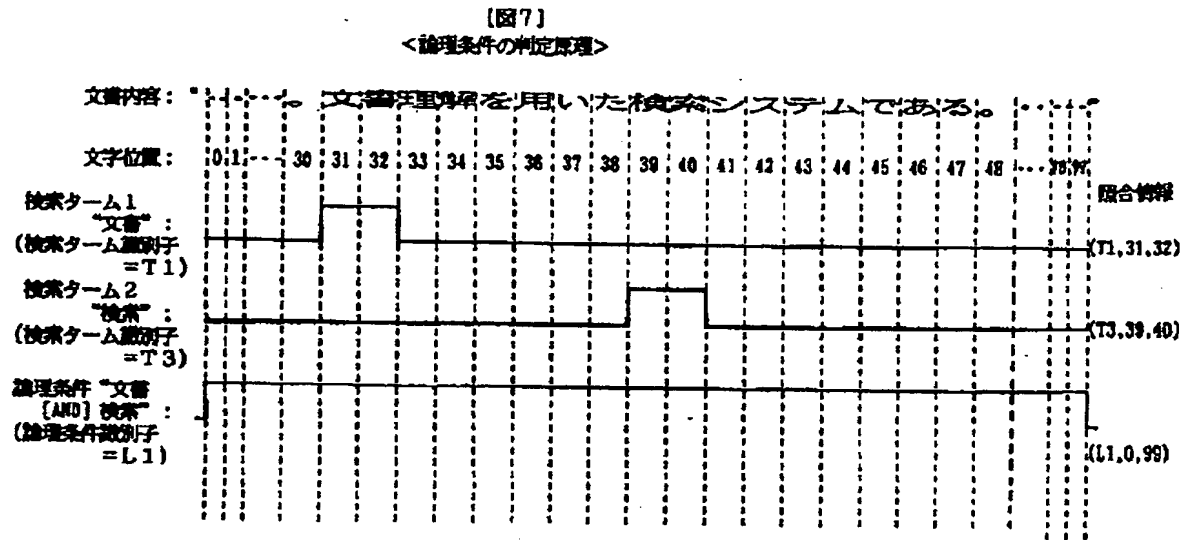


【図6】

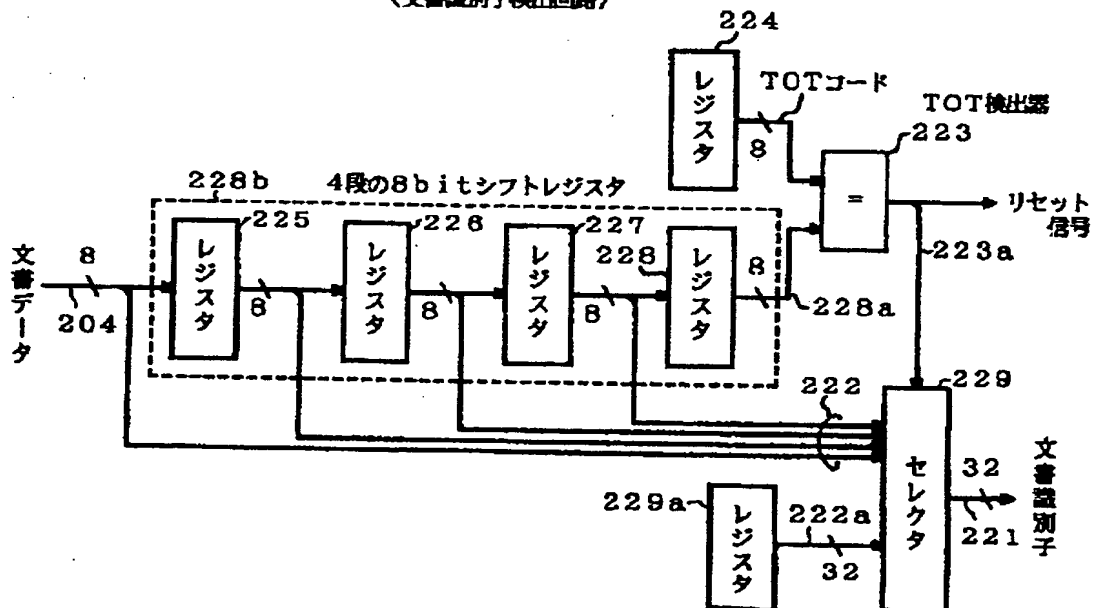
【図6】  
 <文脈条件の判定原理>



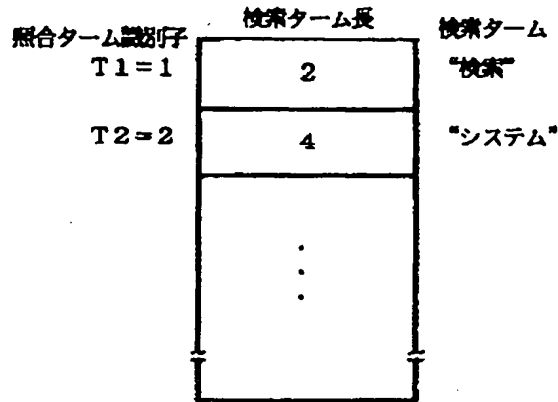
【図7】



【図8】

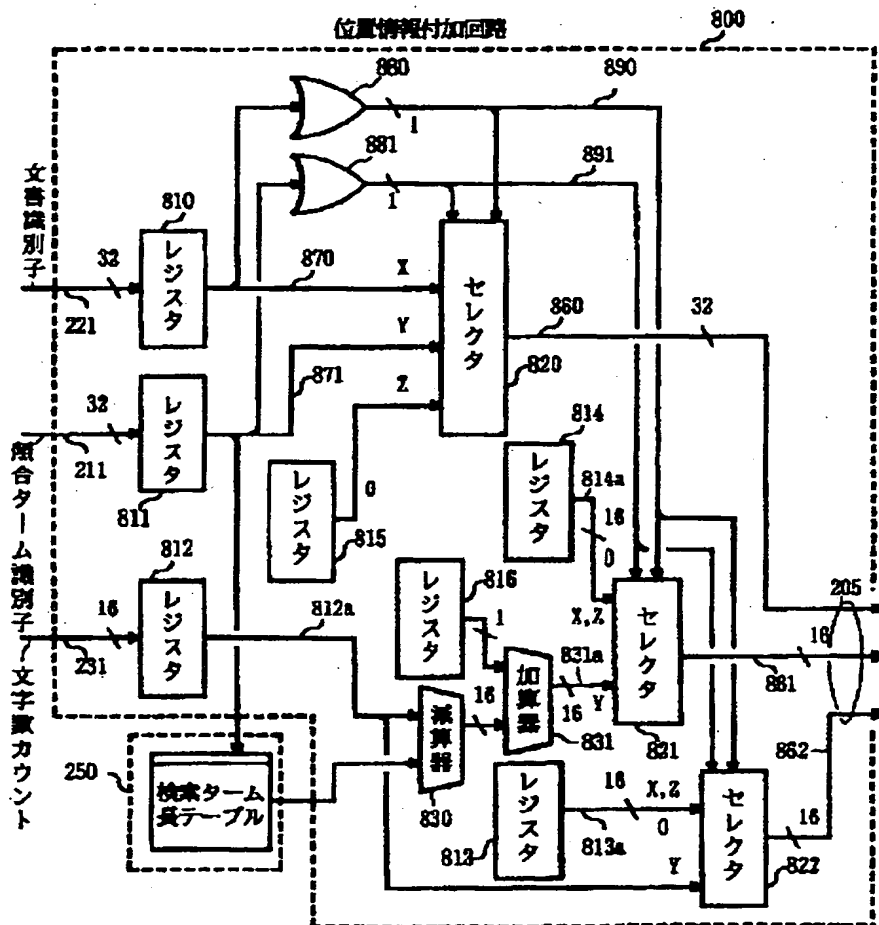
【図8】  
＜文書識別子検出回路＞

【図9】

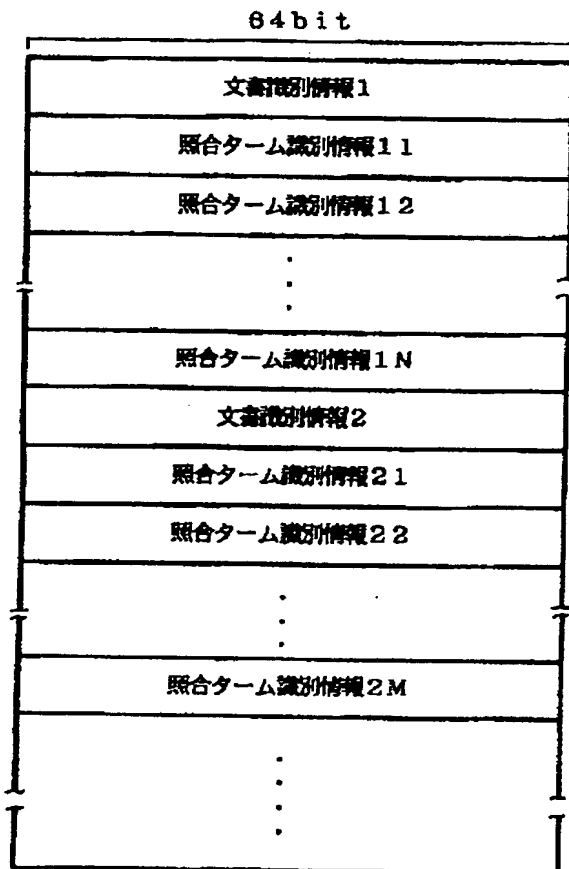
【図9】  
(検索ターム長テーブルの構造)

【図10】

【図10】



【図13】

【図13】  
〈照合情報の構造〉

【図24a】

【図24a】  
〈論理条件の処理ステップ(その1)〉

初期状態: バッファ370

(D1, 0, 0)
(T1, 31, 32)
(T2, 33, 34)
(P1, 31, 34)
(T3, 39, 40)
(T4, 41, 44)
(C1, 30, 48)
(C2, 30, 48)

ワークエリア

出力バッファ331
-----------

ステップ1: バッファ370

(T1, 31, 32)
(T2, 33, 34)

ワークエリア

(D1, 0, 0)
出力バッファ331
(D1, 0, 0)

【図24b】

【図24b】

ステップ2: バッファ370

(T2, 33, 34)
(P1, 31, 34)

ワークエリア

(T1, 31, 32)
出力バッファ331
(D1, 0, 0)
(T1, 31, 32)

ステップ8: バッファ370

(C1, 30, 48)
(C2, 30, 48)

ワークエリア

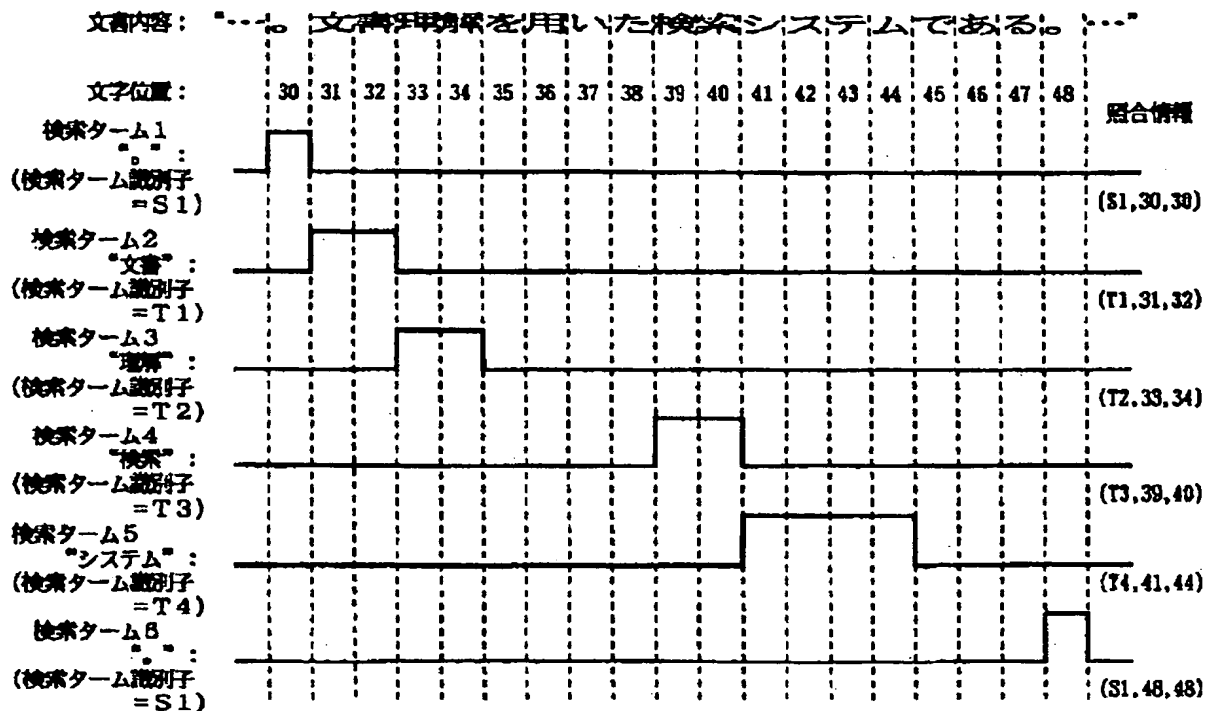
(T4, 41, 44)
出力バッファ331
(D1, 0, 0)
(T1, 31, 32)
(T2, 33, 34)
(P1, 31, 34)
(T3, 39, 40)
(T4, 41, 44)



【図14】

【図14】

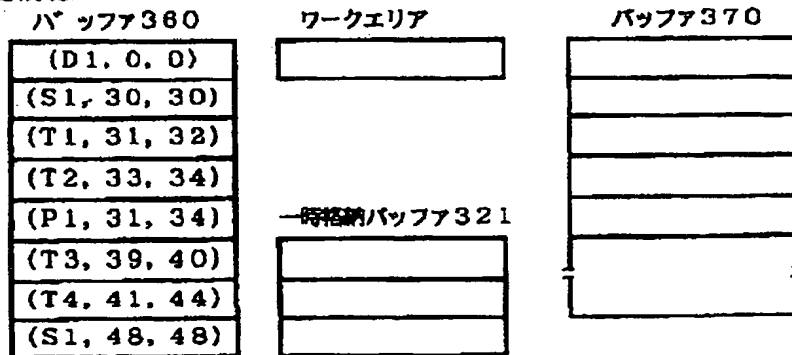
〈文字列照合回路200の出力例〉



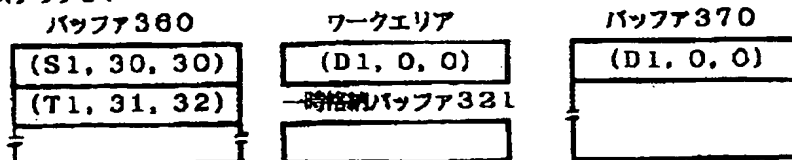
【図20a】

【図20a】〈文庫条件の処理ステップ(その1)〉

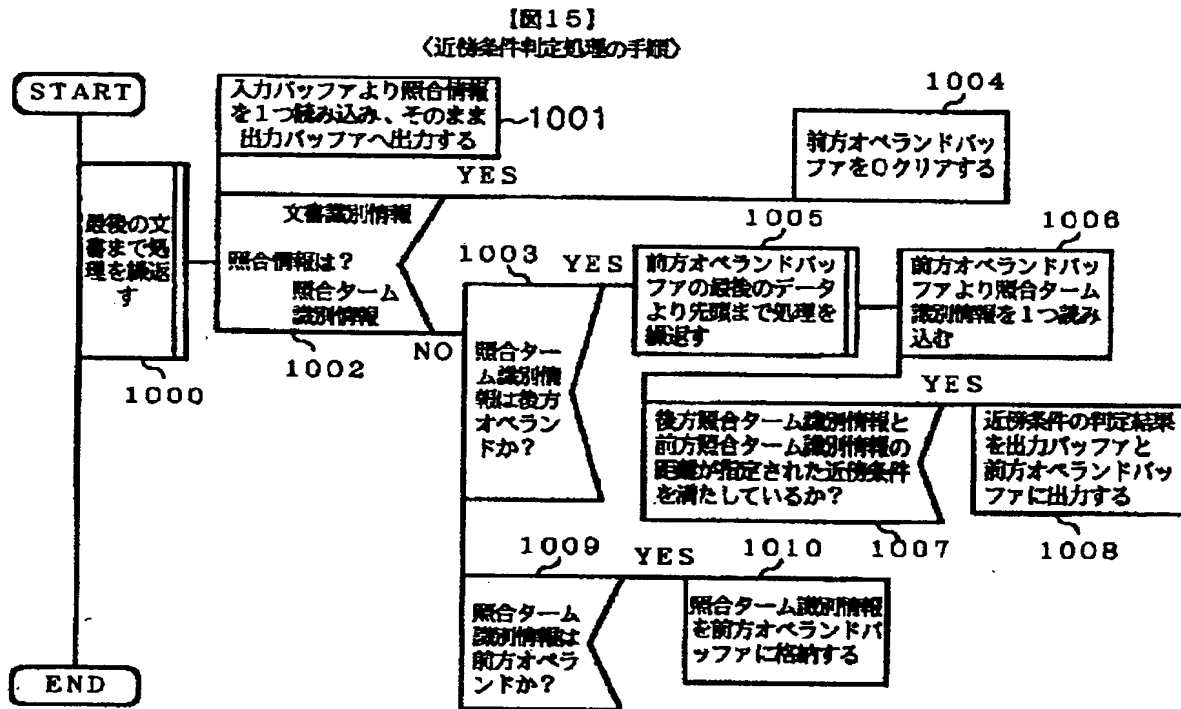
初期状態:



ステップ1:



【図15】



【図16a】

【図16a】  
〈近傍条件の処理ステップ〉

初期状態: バッファ350

(D1, 0, 0)	(8-1)
(S1, 30, 30)	(8-2)
(T1, 31, 32)	(8-3)
(T2, 33, 34)	(8-4)
(T3, 39, 40)	(8-5)
(T4, 41, 44)	(8-6)
(S1, 48, 48)	(8-7)

ワークエリア

--

バッファ360


前方オペランドバッファ311


ステップ1: バッファ350

(S1, 30, 30)	(8-2)
(T1, 31, 32)	(8-3)

ワークエリア

(D1, 0, 0)
------------

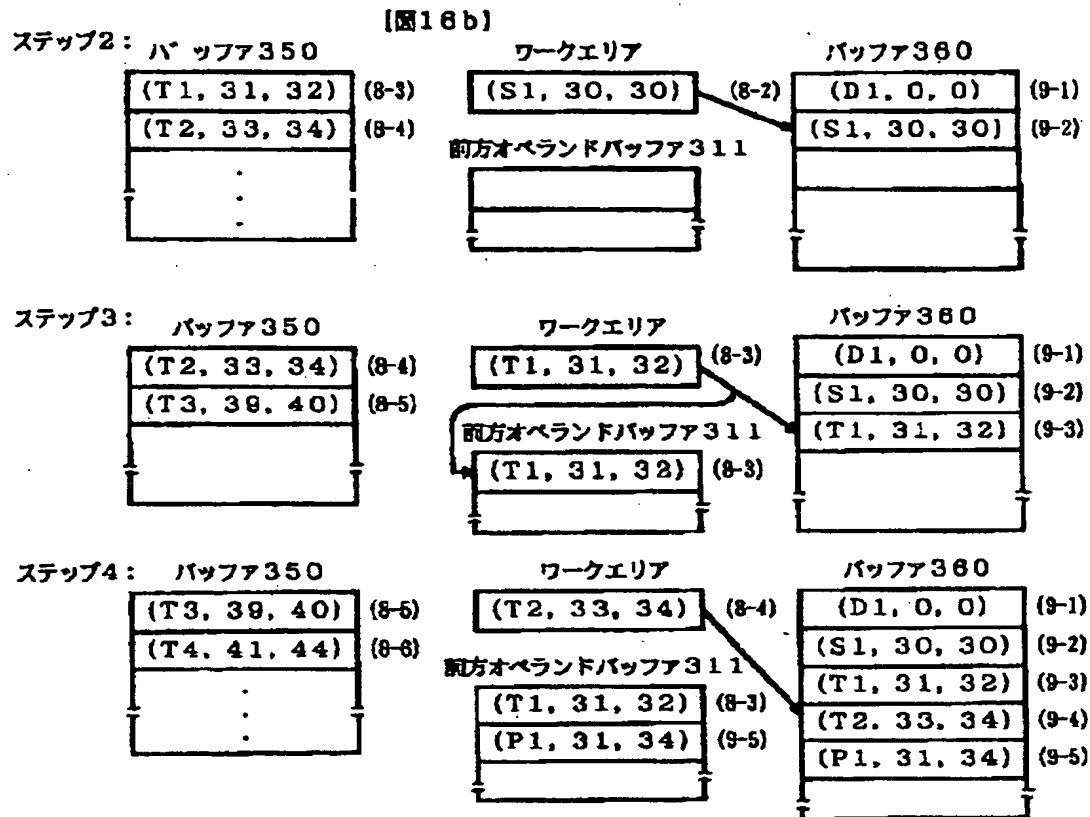
(8-1)

バッファ360

(D1, 0, 0)	(9-1)

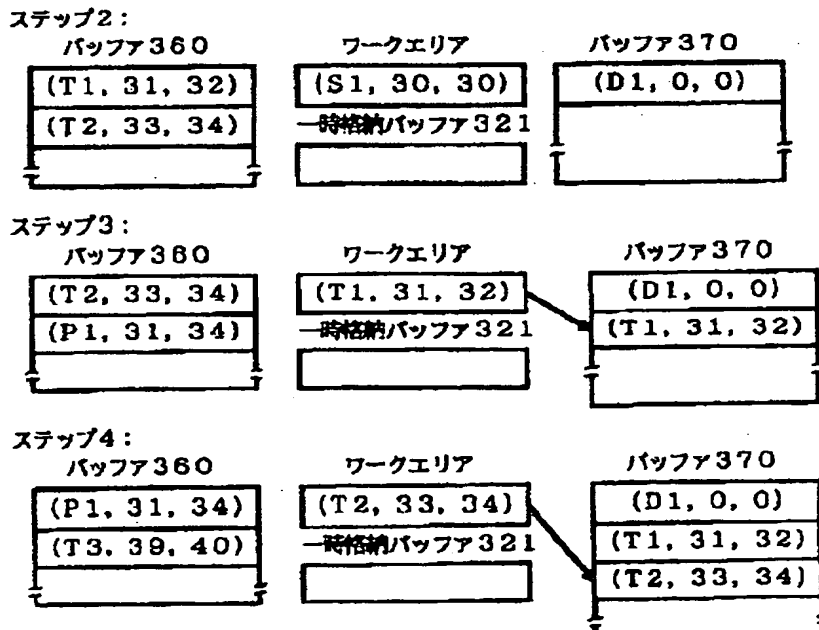
前方オペランドバッファ311


【図16b】



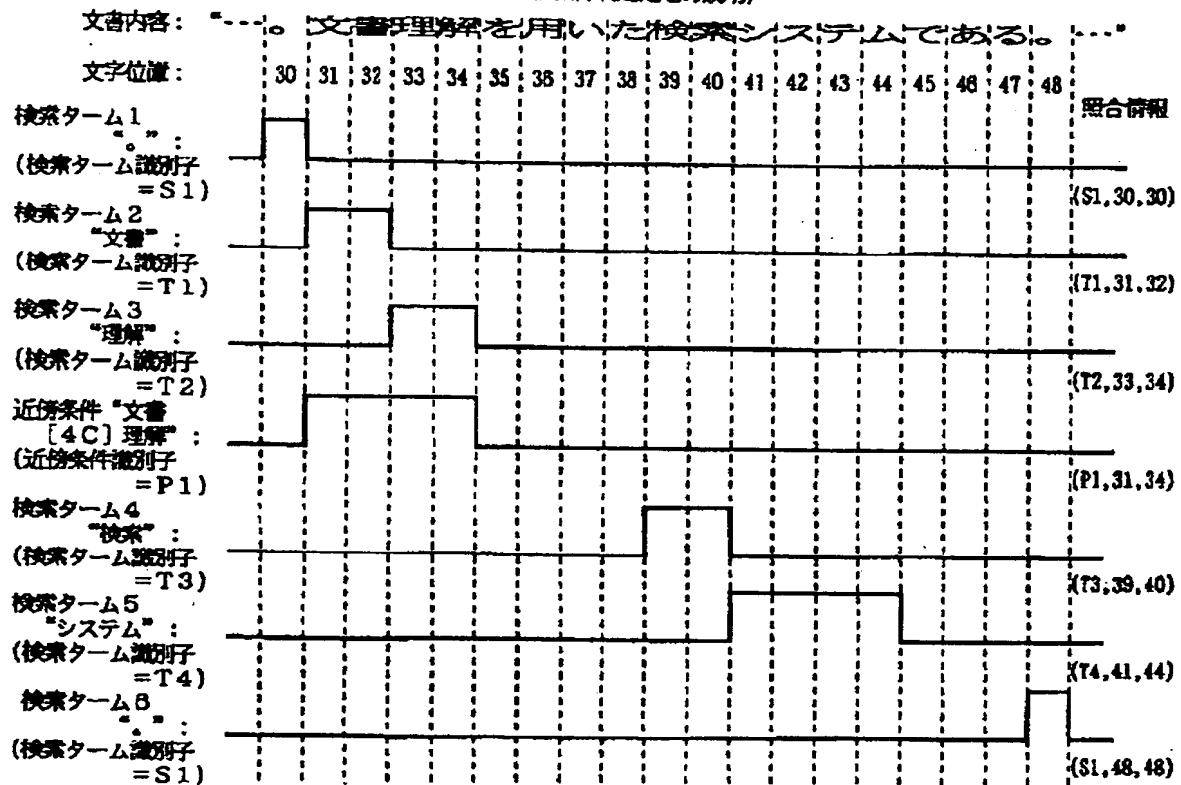
【図20b】

【図20b】



【図17】

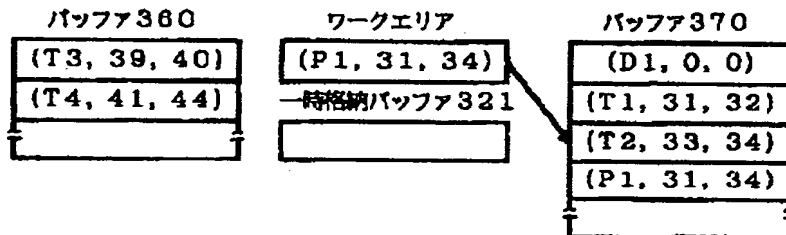
【図17】〈近傍条件判定処理の説明〉



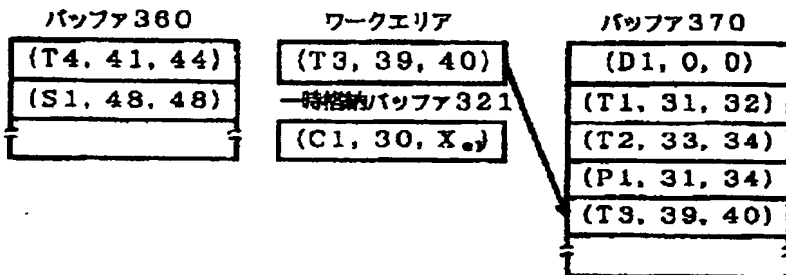
【図21a】

【図21a】〈文脈条件の処理ステップ(その2)〉

ステップ5:

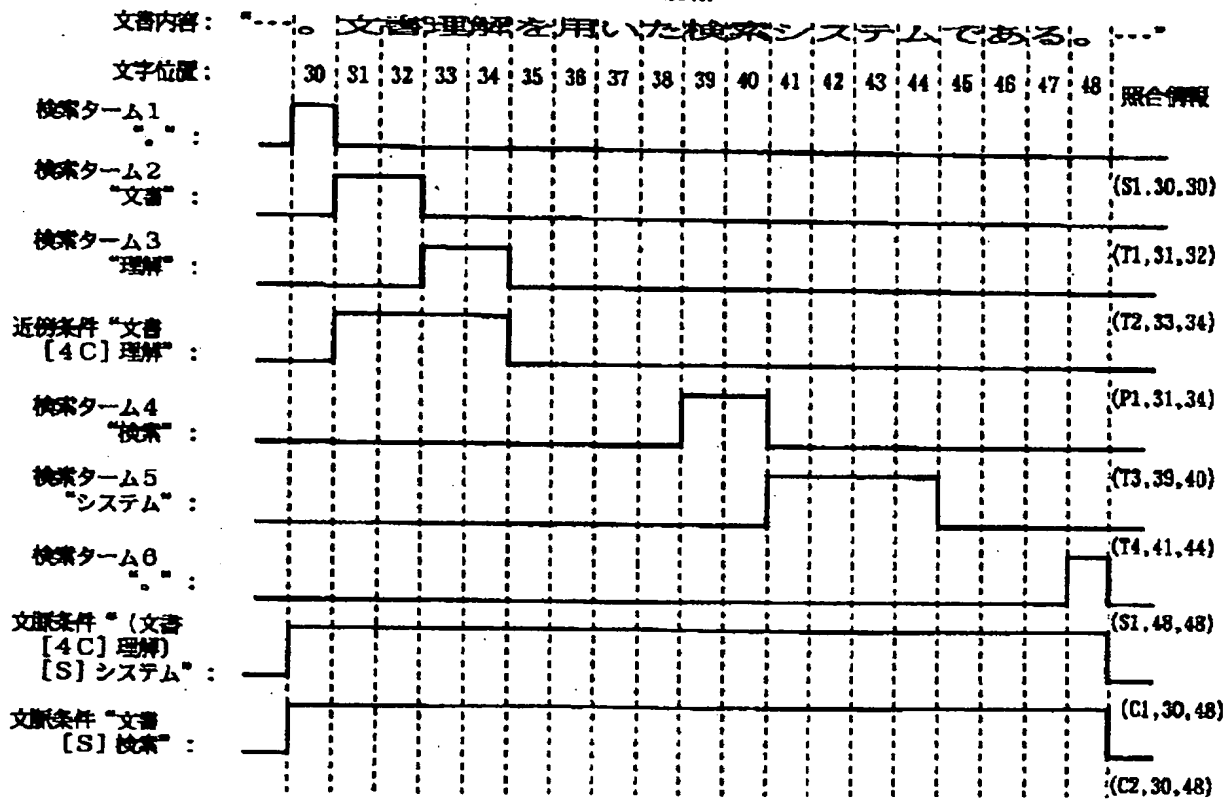


ステップ6:



【図18】

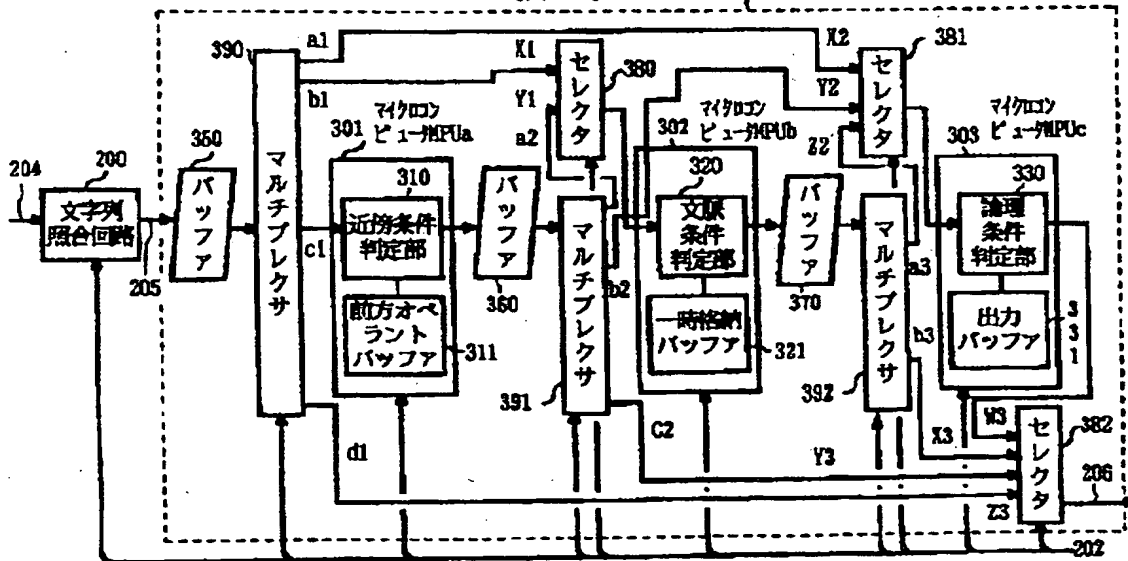
【図18】〈文脈条件の説明〉



【図26】

【図26】

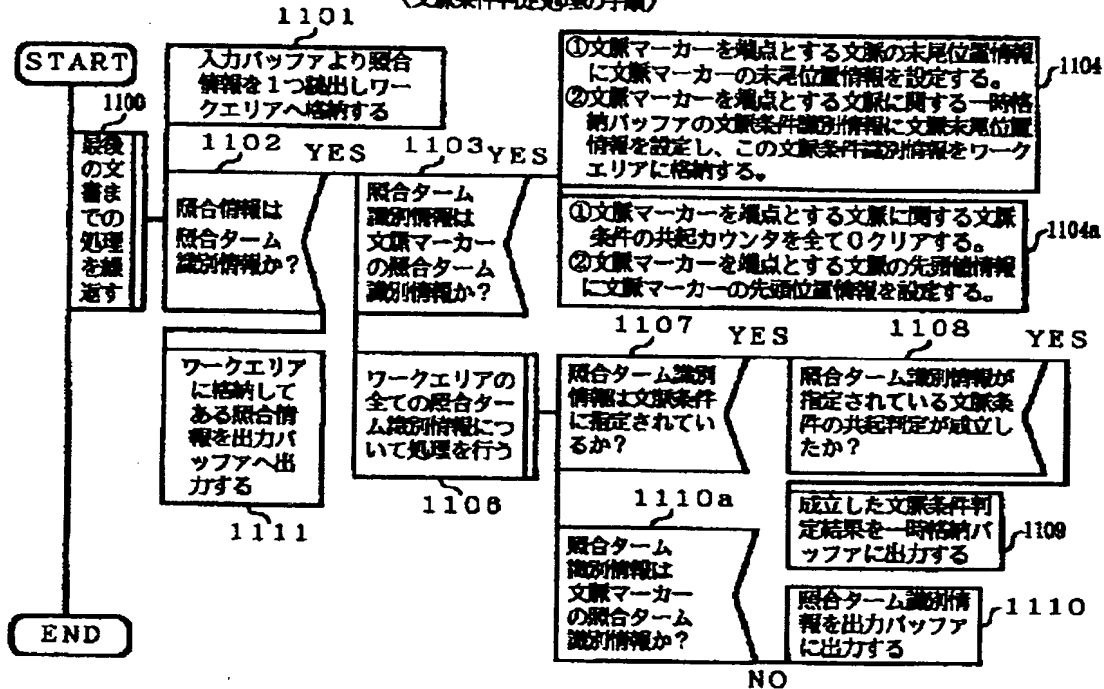
複合条件判定回路300a



【図19】

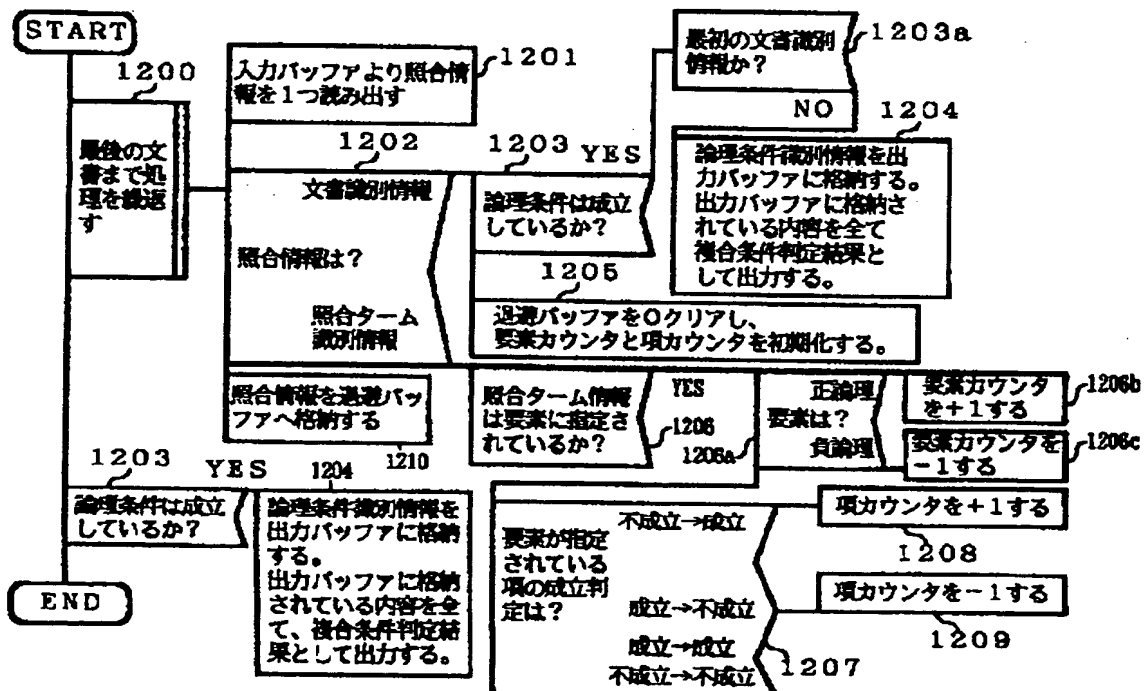
【図19】

〈文脈条件判定処理の手順〉



【図23】

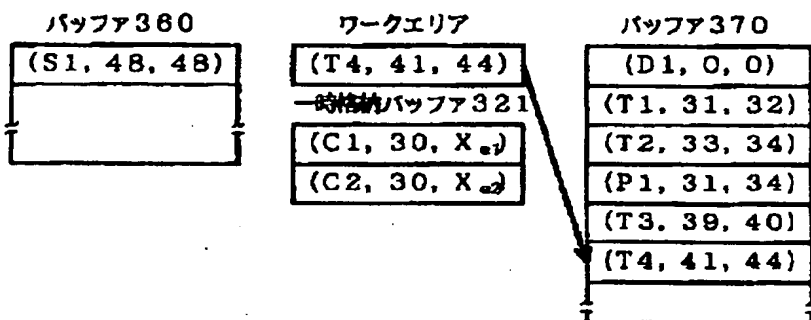
〈論理条件判定処理の手順〉



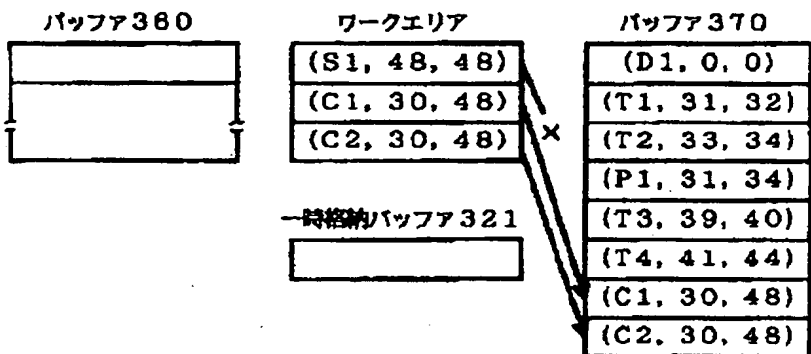
【図21b】

【図21b】

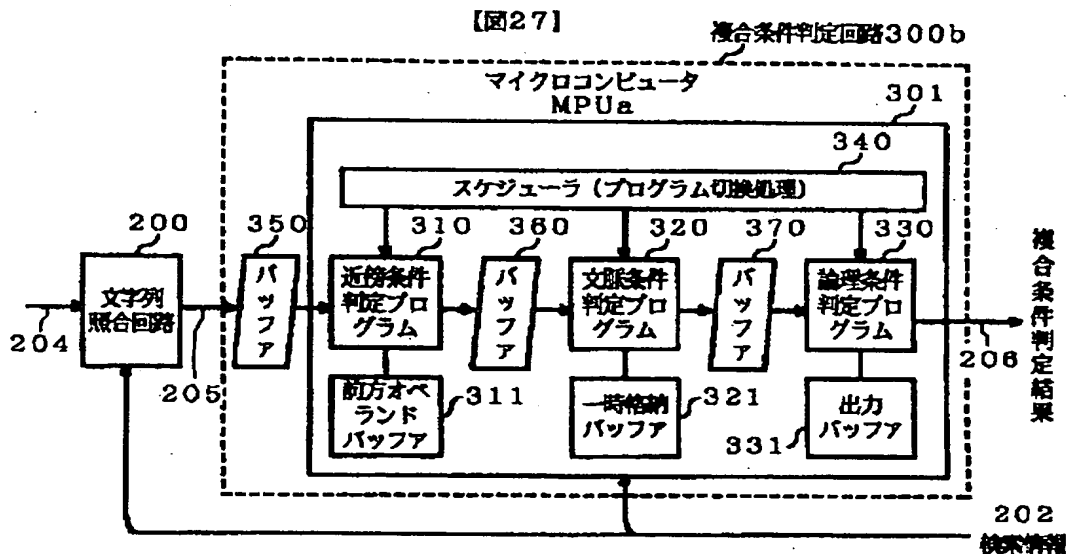
ステップ7:



ステップ8:

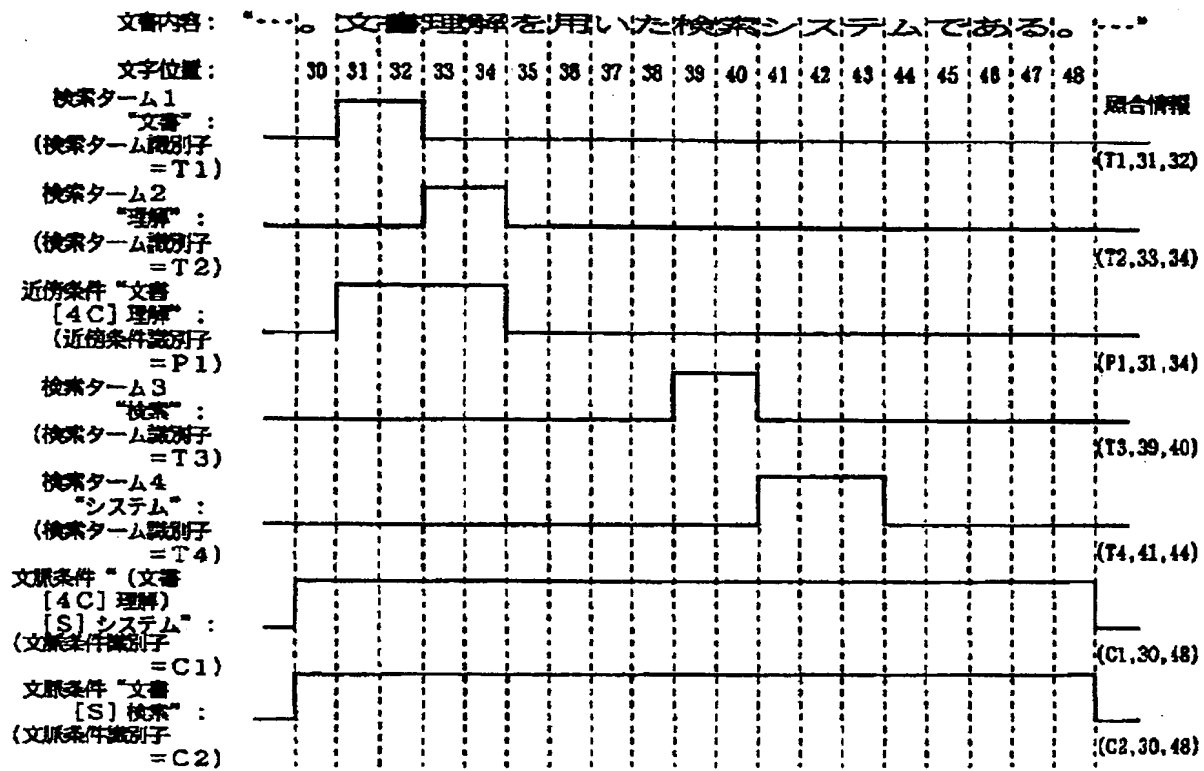


【図27】



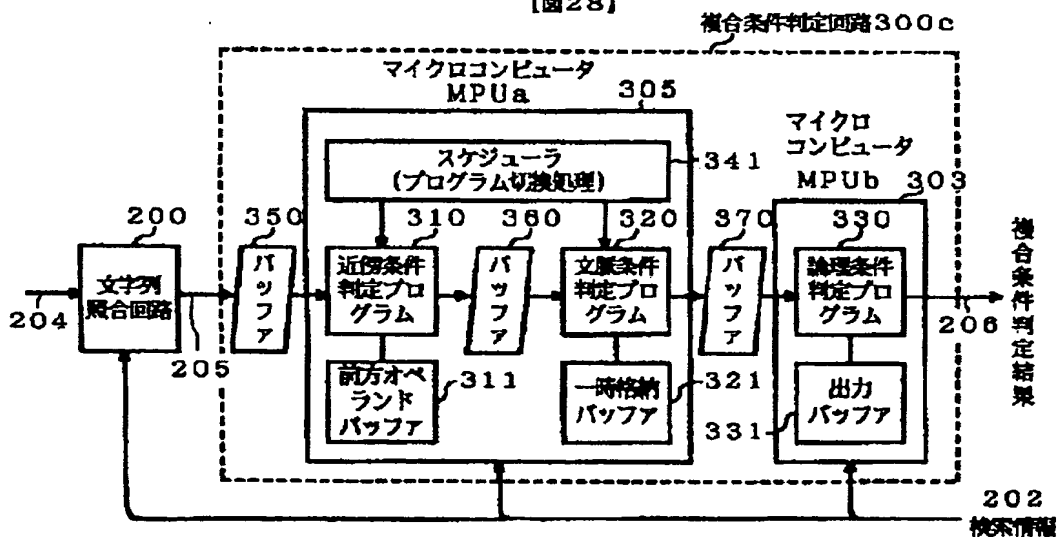
【図22】

【図22】〈文脈条件判定処理の出力例〉



【図28】

【図28】





【図25】

【図25】〈論理条件の処理ステップ(その2)〉

ステップ7:

バッファ370

(C2, 30, 48)

ワークエリア

(C1, 30, 48)
出力バッファ331
(D1, 0, 0)
(T1, 31, 32)
(T2, 33, 34)
(P1, 31, 34)
(T3, 39, 40)
(T4, 41, 44)
(C1, 30, 48)

ステップ8:

バッファ370

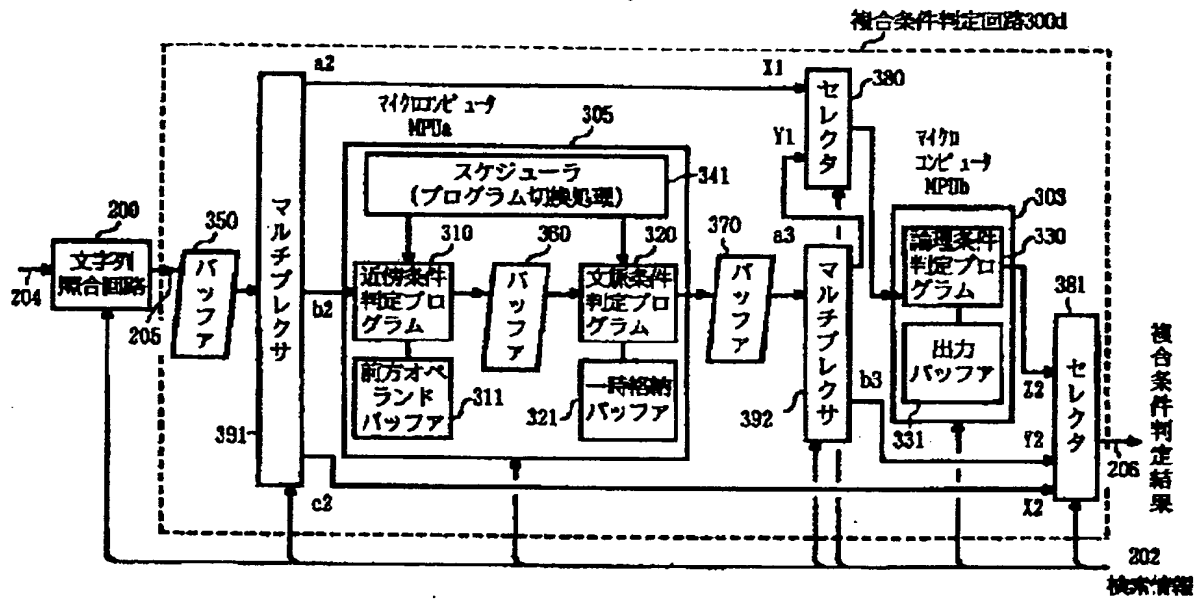

ワークエリア

(C2, 30, 48)
出力バッファ331
(D1, 0, 0)
(T1, 31, 32)
(T2, 33, 34)
(P1, 31, 34)
(T3, 39, 40)
(T4, 41, 44)
(C1, 30, 48)
(C2, 30, 48)
(L1, 0, 99)

場合条件判定  
結果208と  
して出力する

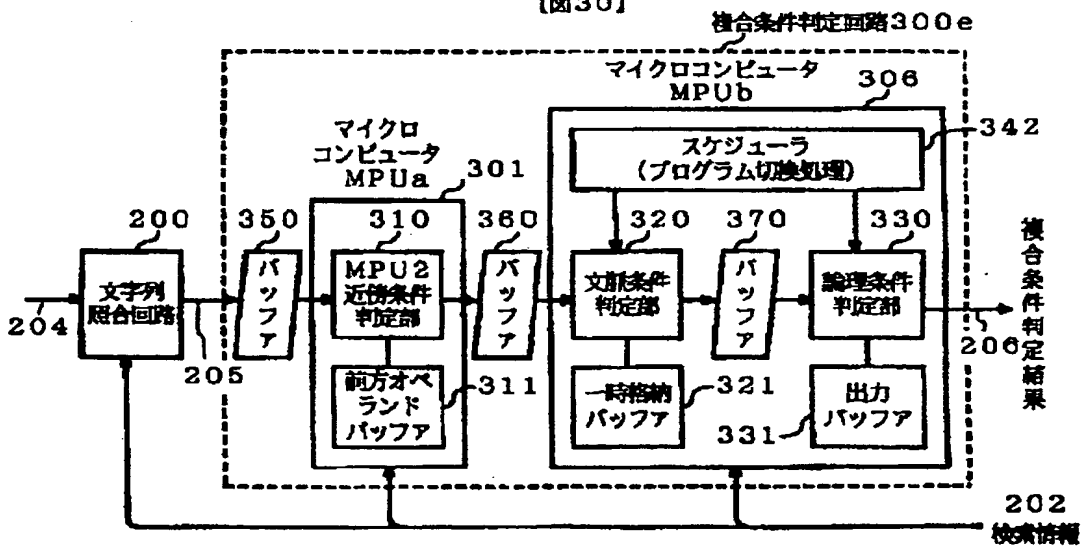
【図29】

【図29】



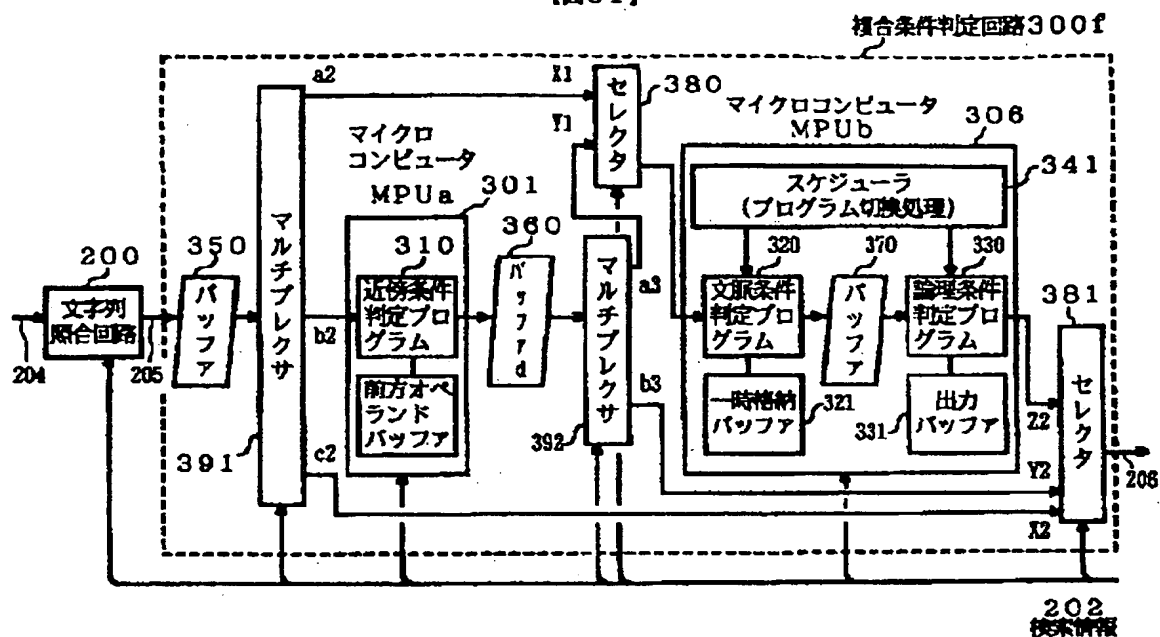
【図30】

【図30】



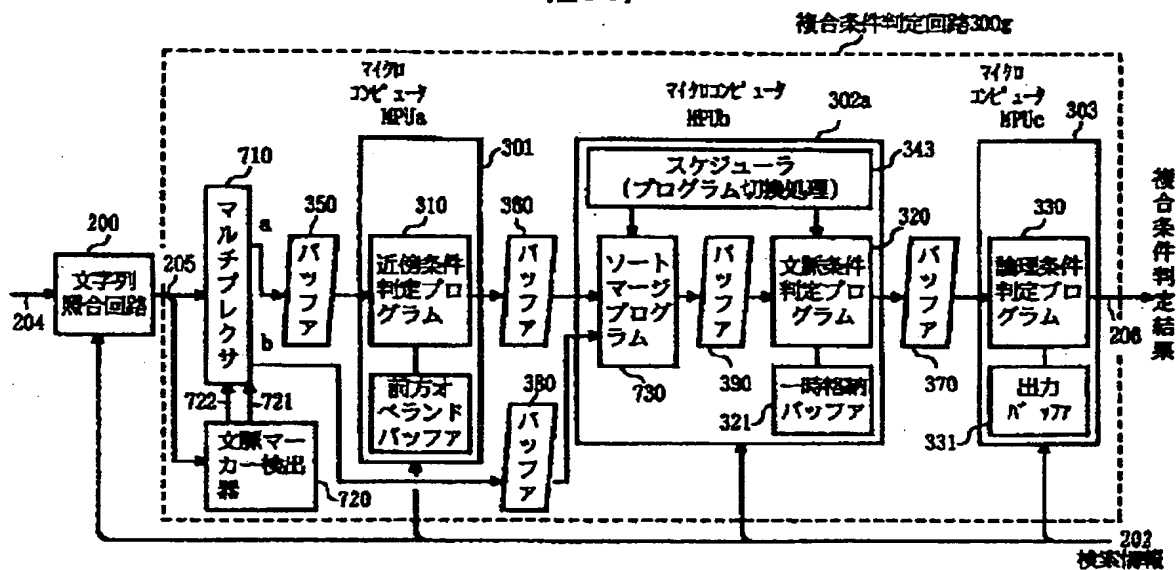
【図31】

【図31】



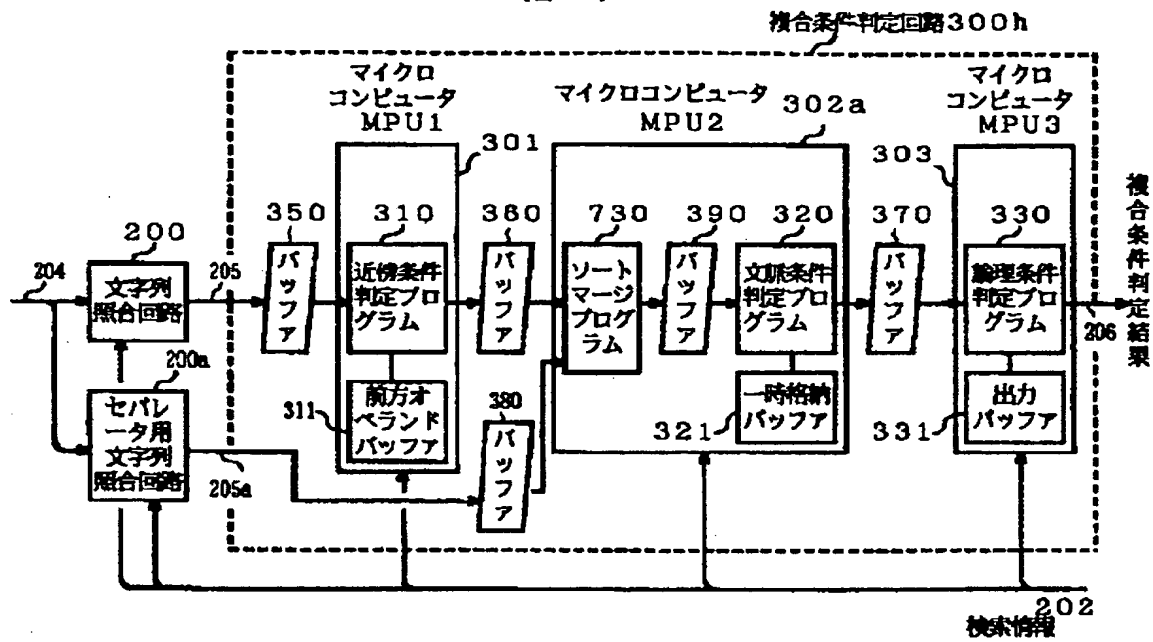
【図32】

【図32】



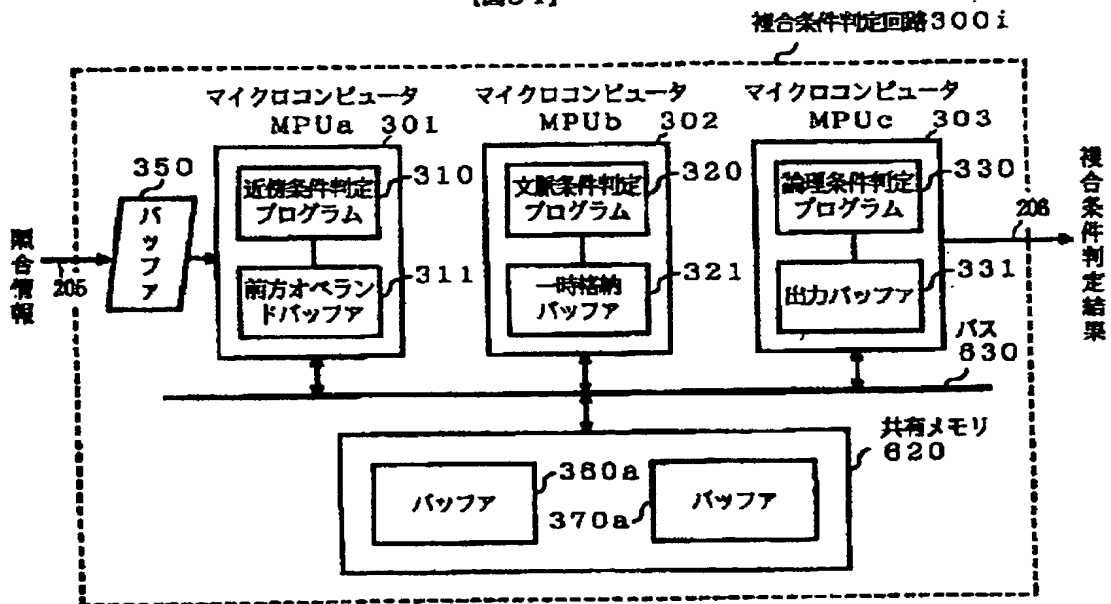
【図33】

【図33】

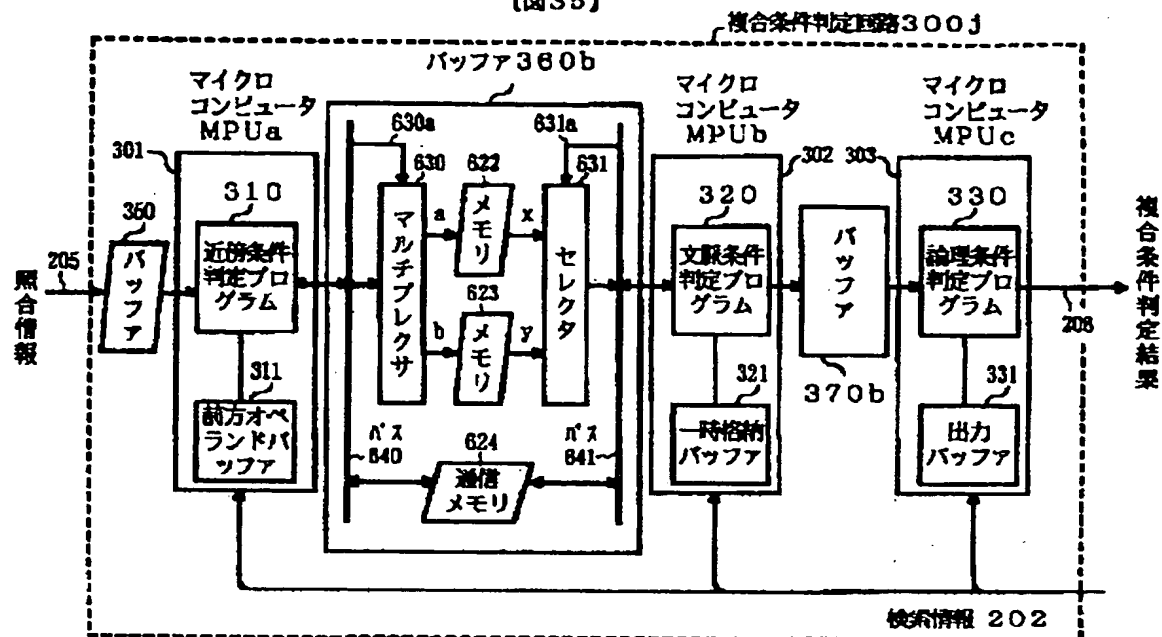


【図34】

【図34】



【図35】



フロントページの続き

(72)発明者 畠山 敦  
 東京都国分寺市東恋ヶ窪1丁目280番地  
 株式会社日立製作所中央研究所内

(72)発明者 藤澤 浩道  
 東京都国分寺市東恋ヶ窪1丁目280番地  
 株式会社日立製作所中央研究所内

**THIS PAGE BLANK (USPTO)**